

Dones de categoria: anàlisi del biaix de gènere a les categories de Viquipèdia

Núria Ferran-Ferrer
Miquel Centelles
Yessica Macià
Juan-José Boté-Vericad
Julià Minguillón

Encàrrec de la Xarxa Vives d'universitats. Juny de 2023. *Dones de Categoria: anàlisi del biaix de gènere a les categories de Viquipèdia. Informe de diagnosi tècnica, posicionament acadèmic i proposta de millora del sistema d'organització del coneixement de Viquipèdia.*

Facultat d'Informació i Mitjans Audiovisuals
Universitat de Barcelona
Melcior de Palau, 140
08014 Barcelona

Xarxa Vives d'Universitats
Universitat Jaume I
Edifici Àgora, Local 10
12006 Castelló de la Plana

Women and Wikipedia: Analysis of the Gender Gap in the Co-Production of Knowledge in the Wikipedias of Spain (W&W), projecte finançat pel Plan Nacional I+D+I del Ministerio de Ciencia e Innovación (MCIN) i la Agencia Española de Investigación (AEI) (Reference Grant PID2020-116936RA-I00).



Autoria: Núria Ferran-Ferrer, Miquel Centelles, Yessica Macià, Juan-José Boté-Vericad i Julià Minguillón

Obra distribuïda sota una llicència Creative Commons BY-NC-SA 4.0



Cita recomanada

Ferran-Ferrer, Núria; Centelles, Miquel; Macià, Yessica; Juan-José Boté-Vericad; Minguillón, Julià (2024). *Dones de categoria: anàlisi del biaix de gènere a les categories de Viquipèdia*. Xarxa Vives d'Universitats; Centre de Recerca en Informació, Comunicació i Cultura (CRICC), Universitat de Barcelona. <http://hdl.handle.net/2445/208281>

Sobre les persones autores

Núria Ferran-Ferrer és professora agregada a la Facultat d'Informació i Mitjans Audiovisuals a la Universitat de Barcelona on dirigeix el programa de Doctorat en Informació i Comunicació. És la delegada del rector per la Direcció de la Unitat d'Igualtat. Té el doctorat Europeu per la Universitat de Barcelona (UB, 2010), llicenciada en Periodisme (UAB, 1998), en Documentació (UOC, 2003) i màster en Societat de la Informació (IN3-UOC, 2005). Ha estat professora associada en diferents universitats: UAB, UB i UPF i ha fet estades a la Universitat de Sheffield i a la Universitat de Tallin. És investigadora principal del projecte Women and Wikipedia (W&W). De forma permanent intenta col·laborar amb associacions i entitats culturals amb qui comparteix aspiracions per realitzar canvis socials. De fet, entén la recerca com una eina al servei de la societat i amb impacte sobre les seves necessitats.

ORCID: 0000-0002-9037-8837

Contacte: nferranf@ub.edu

Miquel Centelles és professor a la Facultat d'Informació i Mitjans Audiovisuals de la Universitat de Barcelona. Graduat en Biblioteconomia i Documentació i llicenciat en Filologia. La seva tasca docent i investigadora se centra en la representació i organització de la informació, així com en l'aplicació de tecnologies semàntiques per a la gestió de la informació i el coneixement. Va coordinar el programa de màster en Gestió de Continguts Digitals del 2005 al 2008, i des del 2020 coordina el màster en Humanitats Digitals, que involucra cinc facultats de la UB. Ha col·laborat en diferents projectes de recerca com el projecte Archiver per la preservació de dades de recerca digital (Archiver TENDER - European Union) i també en el disseny de la metodologia Sapiens per part d'elBullifoundation, que dirigeix Ferran Adrià. Ha fet estades internacionals com la més recent a la Universitat d'Illinois. És investigador del projecte Women and Wikipedia.

ORCID: 0000-0003-1739-4889

Contacte: miquel.centelles@ub.edu

Yessica Macià és doctoranda del programa de doctorat d'Informació i Comunicació de la Facultat d'Informació i Mitjans Audiovisuals (FIMA) de la Universitat de Barcelona (UB) des del 2023. És investigadora del projecte Women and Wikipedia.

ORCID: 0000-0002-5509-9656

Contacte: ymaciama7@alumnes.ub.edu

Juan-José Boté és professor de la FIMA de la Universitat de Barcelona (UB). Doctor en Filosofia (Lingüística i Ciències de la Informació) per la Universität Hildesheim (2022) i doctor en Informació i Documentació en la Societat del Coneixement per la Universitat de Barcelona (2013). És membre de la comissió d'igualtat i de la comissió de qualitat de la facultat. És membre de l'Institut Interuniversitari d'Estudis de Dones i Gènere (IIEDG). Ha estat professor associat a diferents universitats: UAB, VIU, UOC. Ha participat en diferents projectes de recerca com Ciencia abierta en España (RTI2018-094360-B-I00), el projecte Archiver, i ara, Women and Wikipedia.

ORCID: 0000-0001-9815-6190

Contacte: juanjo.botev@ub.edu

Julià Minguillón és professor agregat dels Estudis d'Informàtica, Multimèdia i Telecomunicacions de la UOC des de 2001. Doctorat en Enginyeria Informàtica per la UAB l'any 2002. És investigador del projecte Women and Wikipedia.

ORCID: 0000-0002-0080-846X

Contacte: jminguillona@uoc.edu

Resum executiu

La Wikipedia és el recurs més utilitzat en l'àmbit educatiu, i està entre les webs més utilitzades, al costat de Google o Yahoo, amb més de cinc mil milions de persones que la llegeixen, en les gairebé 300 llengües disponibles, arreu del món. Aquesta enciclopèdia digital ha transformat la forma en què es produeix i es distribueix la informació mitjançant la col·laboració oberta, donat que qualsevol persona pot afegir i editar-ne el contingut. De fet, és mantinguda per una comunitat global de persones voluntàries i, per tant, té el potencial únic de facilitar una producció de coneixement equitativa, basada en béns comuns i la provisió d'espais virtuals de negociació. És sens dubte un dels esforços de cooperació humana més grans que hi ha hagut mai, tant pel que fa al nombre de persones involucrades (centenars de milers) com de magnitud de l'obra que es crea (desenes de milions d'articles). De fet, la Fundació Wikipedia té un objectiu molt ambiciós: "ser la suma de tot el coneixement humà existent".

Tot i aquests assoliments espectaculars, la Wikipedia pateix un biaix de gènere persistent pel que fa al nombre de persones que l'editen, així com de continguts sobre dones. Aquest biaix és conegut des de 2011 i s'han pres mesures, però encara no s'ha aconseguit reequilibrar. Les dades ens indiquen que estem parlant que només un 5% dels continguts de l'enciclopèdia són biografies de dones (el 22% són biografies d'homes)¹ i que menys del 15% de les persones voluntàries que editen Wikipedia són dones.

Aquest biaix de gènere, compartit entre totes les Wikipedies, s'agreuja en algunes enciclopèdies, com el cas de la catalana (la Viquipèdia) amb les categories que classifiquen els continguts de l'enciclopèdia per ajudar a recuperar continguts. La Viquipèdia va acordar en tres negociacions, dins de la seva comunitat de viquipedistes, no utilitzar categories identitàries de gènere, com són "dones" o "persones no-binàries". A la resta de Wikipedies, com la castellana, l'anglesa, o la francesa, els continguts es poden consultar a partir del gènere, ja que permeten la creació de categories a partir de la diferenciació de gènere. Les categories són formes lingüístiques, construccions socials i culturals, que presenten els continguts de l'enciclopèdia. Aquestes categories són un reflex d'una societat que pateix un fort biaix de gènere i no visualitzar la presència dels articles sobre dones o altres gèneres minoritzats no només contribueix a mantenir aquesta clivella, sinó que l'augmenta, i no respon a les necessitats de les persones usuàries que consulten o editen la Viquipèdia.

Per encàrrec del programa d'Igualtat de Gènere de la Xarxa Vives d'Universitats (XVU) presentem l'estudi "Dones de Categoria: Anàlisi del biaix de gènere a les categories de Viquipèdia", un informe de diagnòstic tècnic d'aquesta realitat i que aporta un posicionament acadèmic a la problemàtica a través d'una proposta de millora de la visualització de les diferents identitats de gènere i l'accés als seus continguts a Viquipèdia.

Per a l'elaboració d'aquest informe s'ha fet una revisió de la literatura existent, i s'ha analitzat, d'una banda, la Viquipèdia on s'ha donat el cas de la no acceptació de categories que identifiquin els diferents gèneres. S'han analitzat els arguments emprats durant les tres deliberacions i votacions que hi va haver dins la comunitat per entendre la necessitat de no permetre la visualització de les dones a través d'un recurs de navegació i recuperació de la informació com són les categories. I de l'altra banda, es proposa com a solució utilitzar un altre projecte de Wikimedia, Wikidata, una representació formal del graf de coneixement implícit en el contingut de les pàgines de Wikipedia. Per analitzar tant la Viquipèdia com Wikidata s'ha realitzat una diagnòstic d'aquests dos sistemes d'organització del coneixement a través de tècniques com són el test amb persones usuàries, avaluacions heurístiques, inspecció d'estàndards i entrevistes amb persones expertes. També s'hi estudia la solució que va incorporar la comunitat viquipedista

¹ https://wcd0.wmflabs.org/diversity_over_time/

quan va rebutjar l'ús de categories per gènere a través d'un cercador implementat en algunes pàgines de Viquipèdia. Després de les dues anàlisis, tant de Viquipèdia com de Wikidata, s'estableixen conclusions des d'un punt de vista acadèmic de la presència o absència del biaix de gènere a Viquipèdia i Wikidata.

La solució proposada en aquest informe apunta que Wikidata sigui el sistema d'organització del coneixement de Viquipèdia, ja que Wikidata no augmenta el biaix de gènere present en la societat. Alhora, Wikidata i la seva propietat referida a les identitats de gènere ja etiqueta el 81,93%² de les entitats del tipus ésser humà. A més a més, ja hi ha precedents exitosos que vinculen Wikidata i Viquipèdia com són les infotaules o els registres d'autoritat. Finalment, l'administració de Wikidata compta amb eines que faciliten l'edició de dades i a canvi impulsen la incorporació de la perspectiva de gènere a través de la propietat que recull més diversitat de gènere, visibilitzant identitats, com ara, agènere, dona, home, intersexual, nobinari, transgènere dona i transgènere home, entre d'altres.

Per tant, aquest informe proposa un sistema d'organització del coneixement que adreça dos aspectes cabdals en aquesta problemàtica del biaix de gènere a la Viquipèdia. D'una banda, els aspectes tècnics, que han de permetre recuperar els continguts acord a les necessitats d'informació de les persones usuàries, facilitant, per tant, la visualització de les diferents identitats de gènere, com qualsevol altre aspecte necessari per recuperar la informació. I de l'altra banda, els aspectes socio-culturals, que a través dels arguments acadèmics i tècnics que aporten aquest informe, puguin ajudar en els processos de governança de la comunitat i donar suport a l'evolució de la cultura viquipedista cap a valors més inclusius, que facilitin l'accés i visibilització de continguts i col·lectius minoritzats.

Paraules clau

bretxa de gènere, categories, taxonomies, ontologies, Wikipedia, Viquipèdia, Wikidata

² Dada corresponent a l'1 de juliol de 2023.

Sumari

1.	Introducció	10
2.	Metodologia	14
2.1.	Revisió sistemàtica de la literatura	15
2.2.	Entrevistes a persones expertes	18
2.3.	Ús per part de les persones usuàries	18
2.3.1.	Nombre d'accessos a les categories	18
2.3.2.	Test amb persones usuàries	19
2.4.	Inspecció d'estàndards i guies per a l'avaluació de la taxonomia (Viquipèdia) i les ontologies (Wikidata)	27
2.5.	Proposta d'avaluació heurística de l'esquema de categories de Viquipèdia	28
2.6.	Anàlisi dels processos de presa de decisions	31
3.	Revisió de la literatura	33
3.1.	Wikipedia.....	34
3.2.	Wikidata	36
4.	Objectius	40
5.	Diagnosi de Viquipèdia	42
5.1.	Anàlisi del sistema d'organització del coneixement de Viquipèdia	42
5.1.1.	Categorització del sistema d'organització del coneixement de Viquipèdia	42
5.1.2.	Descripció de l'esquema de categorització	45
5.2.	Avaluació del sistema de categories per a la navegació i cerca.....	53
5.2.1.	Avaluació amb persones usuàries	53
5.2.2.	Avaluació heurística	59
5.3.	Anàlisi del procés de presa de decisions	63
5.3.1.	Anàlisi del procés de presa de deliberació.....	64
5.3.2.	Anàlisi dels arguments de les deliberacions	65
5.3.3.	Anàlisi del procés de votació.....	66
5.4.	Conclusions de la presència o absència de biaix de gènere.....	66
6.	Diagnosi de Wikidata	68
6.1.	Anàlisi del sistema d'organització del coneixement per ontologies	68
6.1.1.	Model de dades.....	68
6.1.2.	Ontologia en el context de Wikidata	72
6.1.2.1.	Propietats.....	73
6.1.2.2.	Classes i instàncies	77
6.1.2.3.	Etiquetatge d'elements i propietats	80
6.1.3.	El fons de dades de Wikidata	81
6.2.	Avaluació del sistema d'ontologies per a la navegació i cerca	83
6.2.1.	Avaluació amb persones usuàries	83

6.2.1.1. Eficiència	83
6.2.1.2. Efectivitat	84
6.2.1.3. Facilitat d'aprenentatge	84
6.2.1.4. Trobabilitat	84
6.2.1.5. Satisfacció.....	85
6.2.2. Inspecció d'estàndards i avaluació heurística	86
6.2.2.1. Instàncies utilitzades com a classes	86
6.2.2.2. Desgavell en els nivells superiors de l'ontologia	86
6.2.2.3. Desviació semàntica	87
6.2.2.4. Cicles o bucles de la propietat subclasse de (P279).....	87
6.2.2.5. Generalització redundat	88
6.2.2.6. Modelatge inconsistent	90
6.2.2.7. Repetició de classes.....	90
6.3. Anàlisi del procés de presa de decisions	90
6.4. Conclusions de la presència o absència de biaix de gènere.....	91
7. Proposta de millora del sistema d'organització del coneixement de Viquipèdia	92
8. Bibliografia	94
9. Agraïments	100
10. Autories	101
11. Glossari	101
Llistat de taules, imatges i figures	103
Annexos	105

1. Introducció

Les tecnologies de la informació i la comunicació semblen haver obert un escenari en el qual el pluralisme i les noves formes d'expressió són més possibles que mai, donant lloc a una esfera pública interconnectada (Benkler, 2006; Castells, 2008; Van Dijk, 2012; Clua et al., 2018). La Wikipedia és part de l'esfera pública actual i té el potencial únic de facilitar una producció de coneixement més equitativa mitjançant la producció col·laborativa basada en béns comuns i la provisió d'espais virtuals per a la discussió. La Wikipedia es defensa com una mena de democràcia intel·lectual, ja que es basa ostensiblement en els principis de deliberació racional, consens i negociació (Black et al., 2008). Per tant, la Wikipedia ha transformat la forma en què es produeix i es distribueix la informació mitjançant la col·laboració oberta (Tkacz, 2014). És, doncs, una enciclopèdia en línia gratuïta i de codi obert que és editada i mantinguda col·laborativament per una comunitat global de persones voluntàries. És sens dubte un dels esforços de cooperació humana més grans que hi ha hagut mai. Tant pel que fa al nombre de persones involucrades com de magnitud de l'obra que es crea. De fet, la Fundació Wikipedia té un objectiu molt ambiciós: "ser la suma de tot el coneixement humà existent".

Va aparèixer a la xarxa l'any 2001 per tal que fos accessible per a qualsevol persona amb connexió a Internet³. Així, qualsevol persona podria llegir-la i també editar-la. La Viquipèdia està disponible en gairebé 300 idiomes, sent el millor resultat en la majoria de cerques de Google, i proporciona informació gratuïta i fiable a més de cinc mil milions de persones diàriament⁴. Val a dir que la Wikipedia en català va ser la segona a aparèixer, amb l'entrada "Àbac" creada la matinada del 17 de març de 2001. Malgrat la popularitat, distribució geogràfica i adopció de Wikipedia, s'han plantejat dubtes sobre el grau de descentralització, flexibilitat i obertura d'aquestes noves oportunitats, ja que es generen dins d'un model econòmic, social i polític preexistent (Hood & Littlejohn, 2018). La Wikipedia té polítiques de contingut com el punt de vista neutral, la no investigació original i la verificabilitat, les quals poden causar dificultats no desitjades per a diversos col·lectius per accedir a l'esfera pública i incloure la seva contribució en la construcció del coneixement a l'enciclopèdia (Ferran-Ferrer et al., 2022). A més a més, donada la seva influència, té el potencial d'amplificar i aprofundir aquestes desigualtats (Acey et al., 2021).

La Viquipèdia pateix un biaix de gènere persistent pel que fa a l'edició i als continguts, reconeguts àmpliament per la literatura acadèmica (Antin et al., 2011; Bear & Collier, 2016; Wagner et al. 2016; Hinnoosaar, 2019). D'una banda, pel que fa a l'edició, el 2015 la Wikipedia va establir l'ambiciós objectiu d'aconseguir una taxa de participació del 25% identificada com a col·laboradores femenines (Leonard & Bond, 2019), però el 2023 aquest objectiu encara no s'ha assolit. Pel que fa a persones editores, en un estudi que van fer part dels autors d'aquest estudi, vam poder determinar el nombre de dones editores de la Viquipèdia en espanyol, que era el 13% de les persones editores (Minguillon et al., 2021). La perspectiva d'anàlisi va ser binària perquè a la Wikipedia només es pregunta pel gènere home, dona o no especifica el gènere, a l'espai de la fitxa de la persona usuària.

D'altra banda, pel que fa al biaix de continguts, només un 5% dels seus articles dedicats a biografies són sobre dones⁵. A més, val la pena assenyalar que els buits d'informació relacionats amb les dones, el gènere i la identitat sexual (Park and Bridges, 2022) no es limiten només a la Viquipèdia en anglès. Fins ara, cap dada ha suggerit que cap de les prop de 300 Wikipèdies estigui lliure de desequilibris de gènere en l'edició i el contingut. La Wikipedia és una comunitat de producció entre iguals que depèn de la participació voluntària de nombroses persones. És

³ <https://en.wikipedia.org/wiki/Wikipedia>

⁴ <https://statista.com/statistics/1259907/wikipedia-website-traffic/>

⁵ https://wcdo.wmflabs.org/diversity_over_time/

compreensible que aquestes persones voluntàries creïn i revisin principalment contingut que s'ajusti als seus interessos personals (Worku et al. 2020). Per tant, entendre els factors que contribueixen a excloure determinats continguts de l'enciclopèdia és una qüestió de gran importància.

Al mateix temps, el sistema de verificabilitat de l'enciclopèdia, que no permet compartir contingut original i que sempre cal referenciar amb una font d'informació primària contrastada⁶, dificulta la creació de continguts sobre dones, ja que en la nostra societat les dones tenen moltes dificultats d'accés a l'esfera pública, amb veu i presència, sobretot als mitjans de comunicació generalistes, la tipologia de font d'informació més utilitzada per fonamentar els articles de Wikipedia (Macharia, 2020; Matud et al., 2021). És el que anomenem l'efecte mirall de Wikipedia respecte de la realitat (Eckert & Steiner, 2013).

Amb tot aquest efecte mirall s'agreuja amb el problema sistèmic que pateix l'enciclopèdia (Ford and Wajcman, 2017). I és que diferents estudis postulen que reduir la bretxa de gènere a la Viquipèdia requerirà més que un augment de la contribució femenina; exigirà una comprensió més profunda per part de la Viquipèdia de la seva pròpia cultura de producció de coneixement (Menking & Erikson, 2015), remuntant-la a les seves arrels fonamentals de cultura enciclopèdica i oberta que pot amagar pràctiques d'exclusió (Ford & Wajcman, 2017). Els estudiosos han argumentat que és imprescindible comprendre la cultura única de producció de coneixement a la Viquipèdia per reconèixer i superar les barreres sistèmiques que impedeixen la participació i les contribucions de les dones a la plataforma. Per a aquestes autories, el problema rau en les lògiques més profundes incrustades en les relacions d'infraestructura del propi projecte tecnocientífic" (Ford & Wajcman, 2017). L'objectiu no és només reduir el biaix, sinó també fer de la Viquipèdia un lloc més robust, fiable i transparent per a la producció de coneixement (Menking & Rosenberg, 2021). Concretament, els acadèmics han demanat un replantejament de la qüestió de gènere a la Viquipèdia com un problema sistèmic que requereix canvis en la cultura de producció de coneixement de la plataforma (Ford & Wajcman, 2017).

Aquest problema pot visibilitzar-se clarament en els espais de deliberació de les biografies, els "Articles nominats per esborrar", on una persona editora amb privilegis, l'administrador, pot proposar biografies per esborrar per diversos motius, habitualment la raó per esborrar els articles de dones és manca de notorietat o bé manca de fonts d'informació que la fonamentin (Tripodi, 2023). La progressió dels articles de dones que s'esborren és molt més gran que els articles que es creen, per tant, si la progressió continua així, d'aquí a uns anys les dones s'extingiran de l'enciclopèdia col·laborativa (Ramírez-Ordóñez & Ferran-Ferrer, 2023).

Un altre espai on es detecta el problema sistèmic és el sistema d'organització del coneixement de la Viquipèdia. Els continguts a la Wikipedia s'organitzen amb categories que responen a les cosmovisions de les persones editores que les creen. Aquest sistema d'organització del coneixement ha estat criticat i posat en dubte des de l'acadèmia (Brückner and Adams, 2015; Minguillón et al., 2017) i específicament a Catalunya pel col·lectiu "Dones amb categoria"⁷, una campanya que reclama que les biografies de dones s'agrupin sota categories denominades amb el femení plural. Així, per exemple, a la Viquipèdia, Gemma Mengual apareix a la categoria 'Nedadors sincronitzats', on tot són dones, excepte una persona que és home, que també forma part d'una categoria que sí que està acceptada i que s'utilitza habitualment com és "Catalans del sud". El mateix passa amb 'Gimnastes rítmics espanyols' o 'Feministes barcelonins', on la majoria dels perfils que s'inclouen són femenins, però això no evita que s'etiqueti amb el masculí genèric. D'aquesta manera, les dones queden ocultes perquè s'han d'incloure en categories masculines. Així mateix, tan sols en comptades ocasions es fa servir l'etiqueta en femení. Seria el

⁶ https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Criteria_d%27admissibilitat

⁷ <https://ca.wikipedia.org/wiki/Viquiprojecte:Categoryzaci%C3%B3/Dones>

cas per exemple de “dones barbudes”. Des del col·lectiu de Dones amb Categoria consideren que només es puguin acceptar categories en masculí és invisibilitzar la presència de les dones, i al segle XXI aquesta idea hauria de ser inacceptable, tal com es va recollir en el manifest “Carta oberta a la comunitat viquipedista catalana”⁸.

Les categories de “dones” i “persones no-binàries” van ser esborrades de la Viquipèdia i la Wikipedia Italiana. La Wikipedia francesa té pendent de fer una votació sobre aquest tema durant aquest any⁹. En el cas de la catalana, es van produir dues deliberacions que van consensuar prohibir l'ús d'aquestes categories, un debat va ser el 2014 i un altre que es va allargar dos anys 2015-2017, en el cas de persones no-binàries el debat va ser el 2018 i va durar un dia. En aquests debats hi van participar 23 persones en total, però només van poder votar 21 persones (de 28 que van votar) editores que complien els requisits d'un mínim de 30 dies d'antiguitat i més de 100 contribucions en l'espai principal (articles) durant els darrers tres mesos en data de 21 de febrer del 2018.

Aquesta decisió que va adoptar la comunitat pot provocar una distribució editorial desequilibrada i una representació esbiaixada del coneixement, que condueix a una representació unilateral dels esdeveniments històrics i dels problemes socials a la Viquipèdia. Això reforça i perpetua la desigualtat de gènere, reduint encara més les oportunitats i la representació de les dones (Ramírez, 2021). La classificació en categories facilita la recuperació de perfils concrets i al mateix temps que posa en relleu les biografies de dones existents a la plataforma.

En canvi, en d'altres Wikipedies, com ara la francesa o l'anglesa, les etiquetes femenines hi són presents des de fa diversos anys. Així, en aquests casos sí que s'hi poden trobar categories com ‘femme archéologue’ o ‘British women archaeologists’, respectivament, per referir-se, en aquest cas, a les dones arqueòlogues valorant així el treball i els èxits de les dones (Graells-Garrido et al., 2015). La importància de representar les diferents identitats de gènere en les diverses Wikipedies és un aspecte crític, i per això s'utilitzen categories en femení per facilitar l'accés i la recuperació dels continguts.

I és que la bretxa de gènere a la Wikipedia té conseqüències importants que van més enllà del seu impacte immediat en la representació en línia dels continguts. Una de les conseqüències més crítiques és la perpetuació dels estereotips i biaixos de gènere que existeixen al món offline, que poden provocar més discriminació i desigualtat (Calvo, 2020). La invisibilitat de les dones i la dificultat per trobar la seva informació a la Wikipedia contribueixen a l'exclusió de les dones (Ford & Wajcman, 2017), esborrant les seves contribucions i èxits del coneixement públic. Alguns articles subratllen que és fonamental abordar aquesta bretxa de gènere i garantir que la representació del coneixement de la Wikipedia sigui justa, el màxim d'imparcial i reflecteixi una àmplia gamma de perspectives (Zheng et al., 2022).

Estudis recents s'han focalitzat a treballar amb Wikidata, la base de dades que dona suport a Wikipedia així com d'altres projectes de la Wikimedia Foundation com són Wikicommons. En concret, l'estudi de Zhang and Terveen (2021) descriu que les entrades a Wikidata de dones estan al voltant del 22% (recordem que les biografies de dones són un 5% dels continguts de Wikipedia) i observen que dins d'un conjunt seleccionat de professions on van obtenir dades de distribució de gènere, Wikidata no és més esbiaixada que el món real: homes i dones s'inclouen en percentatges similars, i la qualitat dels elements que representen homes i dones també és similar. Wikidata està mantinguda per una àmplia comunitat d'usuaris de forma col·laborativa

⁸ <https://donesenxarxa.cat/feminismes/carta-oberta-a-la-comunitat-viquipedista-catalana-donesambcategoria>

⁹ https://fr.m.wikipedia.org/wiki/Wikipédia:Sondage/Féminisation_des_catégories

(Piscopo & Simperl, 2018), forma part del projecte Linked Data web (Erxleben et al., 2014), i és multilingüe (Kaffee et al., 2017).

Basant-nos en aquesta comprovació científica que Wikidata no incrementa el biaix de gènere més enllà de l'existent a la realitat i en el marc del projecte Women and Wikipedia (Reference Grant PID2020-116936RA-I00)¹⁰ finançat pel Ministerio de Ciencia e Innovación, que planteja l'anàlisi de la bretxa de gènere a la Wikipedia, s'ha plantejat analitzar, per encàrrec de la Xarxa d'Universitats Lluís Vives, el biaix de gènere al sistema d'organització del coneixement de la Viquipèdia conegut pel nom de "categories" i la possibilitat de millorar la visualització i l'accés a les diferents identitats de gènere de Viquipèdia, a través de les "ontologies" de Wikidata, un sistema d'organització del coneixement que, fins al moment, no se li ha detectat que perjudiqui la presència de les identitats de gènere.

¹⁰ <https://www.ub.edu/wikiwomen>

2. Metodologia

Per realitzar la diagnosi i les propostes de millora del biaix de gènere a les categories de la Viquipèdia s'ha utilitzat una triangulació metodològica, emprant tant una aproximació qualitativa com també la quantitativa. Així, finalment, els mètodes i tècniques de recerca poden sintetitzar-se en la següent taula d'acord amb els objectius que pretenen donar suport per respondre.

Taula 1. Síntesi dels objectius, materials i mètodes del projecte

Objectius	Aproximació metodològica	Mètodes, tècniques i eines
Revisió bibliogràfica	Qualitativa i quantitativa	Revisió sistemàtica de la literatura sobre el biaix de gènere a Wikipedia i a Wikidata: scoping review Atlas.TI i fulls de càlcul
Anàlisi del sistema d'organització de coneixement per categories Avaluació del sistema de categories	Qualitativa i quantitativa	Entrevistes amb persones expertes Anàlisi de logs a partir de dashboard Test amb persones usuàries amb tècnica thinking aloud, recorregut cognitiu i anàlisi amb mètriques: Eficiència Efectivitat Facilitat d'aprenentatge Trobabilitat Satisfacció Caracterització d'estàndards a partir de dimensions d'anàlisi de sistemes d'organització del coneixement estandarditzades i proposades per l'acadèmia i elaboració de proposta d'avaluació heurística i d'anàlisi d'inspecció Anàlisi heurística Indicadors generals: avaluabilitat i reusabilitat Dimensions: nombre de categories i de relacions semàntiques Coordinació entre conceptes Indicadors estructurals: jerarquia, profunditat i amplitud
Anàlisi dels processos de presa de decisions sobre la creació de les categories	Qualitativa i quantitativa	Anàlisi del contingut de les argumentacions de les diferents deliberacions i de la votació de la supressió de categories per gènere

Anàlisi del sistema d'organització del coneixement per ontologies	Qualitativa i quantitativa	Anàlisi del model de dades Anàlisi de les ontologies i el fons de dades de Wikidata Avaluació amb persones usuàries, heurística i inspecció d'estàndards
Anàlisi del procés de presa de decisions a Wikidata sobre el sistema d'organització del coneixement	Qualitativa i quantitativa	Revisió de la literatura
Proposta de millora del sistema d'organització del coneixement de Viquipèdia		

2.1. Revisió sistemàtica de la literatura

Per oferir l'estat de l'art sobre els sistemes d'organització del coneixement a Wikipedia i Wikidata en relació amb les identitats de gènere. Hem emprat el mètode de revisió sistemàtica de la literatura. A continuació, expliquem amb detall com s'ha obtingut un corpus de documents que constitueixen el tema de la nostra anàlisi i com s'han analitzat els resultats. La revisió sistemàtica s'ha dut a terme dur a terme seguint el marc SALSA (vegeu la taula 2) (Grant & Booth, 2009; Booth et al., 2022).

D'una banda es va fer una revisió sistemàtica de la literatura en Wikipedia i gènere per al projecte Women and Wikipedia (article pendent de publicar en una revista i l'informe complet es publicarà a la col·lecció del Centre de Recerca en Informació, Comunicació i Cultura, CRICC). I per altra banda, es va fer una revisió sistemàtica sobre sistemes d'organització del coneixement a la Wikipedia i Wikidata tal com s'explica a continuació.

Taula 2. Adaptat del marc SALSA (Grant & Booth, 2009; Booth et al., 2022) per l'scoping review de Wikipedia i gènere

Step	Criteria
<i>Search</i>	<p><i>Databases: Scopus and Web of Science</i> <i>Period of publication: till December 2022 (included)</i> <i>Searches performed the 10th of January 2023</i></p> <p><i>Scopus - 420</i> <i>Search equation: TITLE-ABS-KEY</i> <i>(wikipedia* AND (gender* OR female* OR male* OR woman OR man OR women OR men OR masculi* OR femini* OR sex OR non-binary OR intersex OR transgen* OR agender OR queer*))</i></p> <p><i>Web of Science - 399. Includes de following databases Web of Science Core Collection, BIOSIS Citation Index, BIOSIS Previews, Current Contents Connect, Derwent Innovations Index, KCI-Korean Journal Database, MEDLINE,, SciELO Citation Index, Zoological Record.</i></p>

	<p>Search equation: <i>TS=TOPIC</i> <i>wikipedia (Topic) and (gender* OR female* OR male* OR woman OR man OR women OR men OR masculi* OR femini* OR sex OR non-binary OR interser OR transgen* OR amended OR queer*) (Topic)</i></p> <p>420 + 399 = 819 330 duplicates between both platforms 488 unique items</p>
AppraisalL	<p>Initial number of documents: 488 Final number after applying inclusion/exclusion criteria: 390 Exclusion non English or Spanish language: 0 Exclusion of Books. We retrieved 33 Books, and we selected just the chapter about Wikipedia and gender: 6 Exclusion of academic articles (comments, letters, editorials, reviews, or papers without method, which does not imply an empirical investigation): 17 Exclusion of the articles out of the period of analysis: 1 paper from 2023 Exclusion of articles when full text not available: 1 (even the author does not have it) Duplications: 8 (communication and paper with same titles. We keep the article) Exclusion during the screening process (title, abstract, keywords) No gender perspective: 31 No WP (Wikidata, or other Wikiverse platforms): 14 WP is not the object of the study (is used as information source or corpus): 313 Total included items: 97 Search of full text (when needed): UB library, academic repositories, Research Gate, personal contact by email with authors</p>
Synthesis	Narrative synthesis based on the analyses carried out
Analysis	<p>Components: Object of study/ Aim: The new knowledge that the article provides around the studied problem is the aim of the research Theme: Readership: about information users of Wikimedia Editing: about contribution and contributors Participation: edithatons, wikiprojects, tea house, etc. Conceptual Framework: Reasons for the gender gap in WP Solutions for the gender gap in WP Further research Limitacions Wikidata and gender</p>

Taula 3. Adaptat del marc SALSA (Grant & Booth, 2009; Booth et al., 2022) per l'scoping review dels sistemes d'organització del coneixement a Wikipedia i Wikidata

Step	Criteria
Search	<p>Databases: Scopus and Web of Science Period of publication: till April 2023 (included) Searches performed the 15 of May 2023</p>

	<p><i>(wikidata OR wikipedia) AND (taxonom* OR ontolog* OR tagg* OR folksonom* OR) AND (gender* OR female* OR male* OR woman OR man OR women OR men OR masculi* OR femini* OR sex OR non-binar* OR intersex* OR transgen* OR agender* OR queer*)</i></p> <p><i>Scopus -20</i> <i>Search equation: TITLE-ABS-KEY</i></p> <p><i>Web of Science - 8.</i> <i>Includes de following databases Web of Science Core Collection, BIOSIS Citation Index, BIOSIS Previews, Current Contents Connect, Derwent Innovations Index, KCI-Korean Journal Database, MEDLINE,, SciELO Citation Index, Zoological Record.</i> <i>Search equation: TS=TOPIC</i></p> <p><i>20 + 8 = 28</i> <i>8 duplicates between both platforms</i> <i>8 unique items</i></p>
<i>AppraisalL</i>	<p><i>Initial number of documents: 20</i> <i>Final number after applying inclusion/exclusion criteria: 11</i></p> <ul style="list-style-type: none"> • <i>Exclusion non English or Spanish language: 0</i> • <i>Exclusion of Books. We retrieved 0 Books, and we selected just the chapter about Wikipedia and gender:</i> • <i>Exclusion of academic articles (comments, letters, editorials, reviews, or papers without method, which does not imply an empirical investigation): 0</i> • <i>Exclusion of the articles out of the period of analysis: 0 paper from 2023</i> • <i>Exclusion articles when full text not available: 0</i> • <i>Duplications: 0 (communication and paper with same titles. We keep the article)</i> • <i>Exclusion during the screening process (title, abstract, keywords)</i> <ul style="list-style-type: none"> • <i>No gender perspective: 0</i> • <i>No WP (Wikidata, or other Wikiverse platforms): 0</i> • <i>WP is not the object of the study (is used as information source or corpus): 9</i> <p><i>Total included items: 11</i> <i>Search of full text: UB library, academic repositories, Research Gate, personal contact with authors</i></p>
<i>Synthesis</i>	<i>Narrative synthesis based on the analyses carried out</i>
<i>Analysis</i>	<p><i>Components:</i> <i>Object of study/ Aim: The new knowledge that the article provides around the studied problem is the aim of the research</i> <i>Results</i></p>

2.2. Entrevistes a persones expertes

Per a l'elaboració d'aquest informe s'ha comptat amb la col·laboració de diferents persones a qui s'ha entrevistat nivell d'expertesa en la matèria. El guió de les entrevistes està accessible a l'Annex.

Taula 4. Persones expertes entrevistades

Participant	Rol	Data
Miriam Redi i Pablo Aragón	Research Team Wikimedia Foundation	27/06/2023, videoconferència
Lydia Pintscher	Lead Product Manager of Wikidata	31/05/2023, videoconferència
Alejandro Morales-Vargas	Facultad de Comunicación, Universidad de Chile.	25/04/2023, videoconferència
Daniele Metilli	Research Fellow at University College London, Department of Information Studies. IP del projecte Wikidata Gender Diversity (WiGeDi)	21/03/2023, videoconferència
Sarai Sanchez Fuentes	Senior UX Designer at Wikimedia Deutschland	10/03/2023, videoconferència

2.3. Ús per part de les persones usuàries

La valoració de l'experiència de les persones usuàries (UX) que fan ús de l'enciclopèdia digital hauria de recollir aspectes relacionats amb la utilitat del lloc, la usabilitat (que sigui fàcil d'utilitzar per part de les persones usuàries), desitjable, trobable, accessible per a qualsevol persona amb la diversitat funcional que sigui, creïble (que sí que pugui confiar) i que aportí valor d'alguna mena a la persona que la faci servir (Morville, 2005). És a dir que l'avaluació per part de les persones usuàries no avalua només si es donen les respostes exactes sinó que també l'experiència d'usar-l'ho sigui agradable i no es limiti només a complir amb un llistat de característiques.

2.3.1. Nombre d'accessos a les categories

Les categories de Viquipèdia no compten amb molt d'ús entre les persones lectores i editors. A tall d'exemple, si agafem les professions feminitzades de Catalunya (Vegeu taula 5) a la Viquipèdia catalana es tradueixen en les categories d'"infirmers", "bibliotecaris" i "mestres" veiem en el dashboard¹¹ que durant el darrer any, els accessos per part de les persones usuàries ha estat al voltant de la trentena de visites i cap variació en l'edició en aquestes tres professions. Si ho comparem en la Wikipedia en castellà, aquestes categories es desdoblen per gènere i tenen la seva categoria en femení. I pel que fa a accessos, la categoria "enfermeros" ha rebut 665 visites i la categoria "enfermeras" 845 en tot l'any 2022.

¹¹ <https://pageviews.wmcloud.org/>

Taula 5. Professions feminitzades a Catalunya

	Àmbit de treball	Ràtio per gèneres
Professions Feminitzades	Personal Infermeria	86,5% dones (dades 2022 Col·legi d'infermeres i infermers ¹² , enfocament recollida dades binari)
	Personal Bibliotecari	82% dones (dades 2022 COBDC ¹³ , enfocament recollida dades binari)
	Personal docent públic Departament Educació	73,5% dones (dades Departament, curs 2022-2023 ¹⁴ , enfocament recollida dades binari)

2.3.2. Test amb persones usuàries

2.2.1.1. Mostratge

Per aquesta part de l'estudi es va comptar amb una mostra de 20 persones, de les quals 9 eren dones, 8 homes i una persona que no ha expressat voluntàriament el seu gènere. 10 eren persones editores (viquipedistes) i 10 lectores. En relació amb la perspectiva de gènere, 10 eren activistes de gènere, membre del projecte Viquidones o de grups de recerca amb perspectiva de gènere, etc. i 10 persones no estaven significades per temes de gènere (Vegeu la taula 6 amb la distribució de la mostra). Les edats dels participants anava dels 22 als 72 anys. La mostra es va reclutar entre l'alumnat i personal de la Facultat d'Informació i Mitjans Audiovisuals (FIMA)¹⁵ de la Universitat de Barcelona, viquipedistes del projecte Viquidones UPF¹⁶, i de la comunitat viquipedista d'Amical Wikimedia¹⁷.

La majoria de les persones usuàries van realitzar el test presencialment a la Universitat de Barcelona (UB) o bé durant una Viquimarató al Museu Nacional d'Art de Catalunya (MNAC). Amb algunes (5 en total) va caldre fer el test en remot, i es va realitzar amb el recurs Teams de Microsoft perquè és el que ofereix la UB al seu personal investigador.

2.2.1.2. Execució dels tests

En tots casos, els tests es va realitzar amb una persona experta en UX del projecte W&W¹⁸ que realitzava l'experiment, prenia notes de la tècnica thinking aloud o pensament manifest (compartir els pensaments en veu alta sobre aspectes positius o negatius del que s'anaven trobant amb la tasca), i s'enregistrava oralment la sessió. L'enregistrament servia per poder calcular la ràtio d'èxit i, qualitativament, interessava per conèixer les opinions i valoracions, que les persones usuàries van donar a conèixer també al qüestionari posterior a la realització dels tests, on valora en del 0 al 5 la satisfacció de cadascuna de les eines tant com la intuïtivitat Dels cercadors i l'explicació de les seves opinions. Totes les persones participants van signar el full de consentiment informat proporcionat per la UB. La duració de cada test va ser d'entre 45 i 75

¹² <https://coib.cat>

¹³ <https://cobdc.org/>

¹⁴ <https://educacio.gencat.cat/ca/departament/estadistiques/personal-administracio-educativa/>

¹⁵ <https://www.ub.edu/portal/web/informacio-mitjans-audiovisuals>

¹⁶ https://ca.wikipedia.org/wiki/Viquiprojecte:Viquidones/Espai_Viquidones_UPF

¹⁷ <https://wikimedia.cat/>

¹⁸ <https://www.ub.edu/wikiwomen/>

minuts i es van realitzar durant els mesos d'abril i maig de 2023. Cap dels participants va rebre una remuneració o compensació per la seva participació.

Taula 6. Distribució per característiques essencials de la mostra emprada pels tests amb 20 persones usuàries

Gènere			Tipologia de persona usuària		Posicionament respecte al gènere	
Dona	Home	No específica	Lectora	Editora	Activista	No posicionat
9	8	1	10	10	10	10
20						

2.2.1.3. Experiments

El guió estava estructurat en 9 tasques (amb una tasca addicional, per les persones editores de la Wikidata) tal com es pot veure en la taula 7 i en més detall, amb el guió complet, a l'Annex B. La totalitat de les 10 tasques cobria tant l'avaluació de la navegació (tasques 6,7 i 10) com la cerca (1-5 i 8-9).

Els diferents experiments o tasques es van establir a partir dels biaixos de gènere en les professions a Catalunya (Vegeu taula 5). En el cas de la primera tasca, “les dones a l'educació”, es va considerar com a professió feminitzada, considerant la majoria de dones professores i mestres com punt de referència per donar possibilitat a la persona usuària de trobar el perfil que li demanaven. En el cas de la segona tasca “de les dones a l'àmbit STEM” (ciència, tecnologia, enginyeria i matemàtiques), es va demanar en els casos on la persona usuària tenia més possibilitats d'especificar la cerca, i per això vam triar un àmbit professionals masculinitzat, en els quals les dones són minoria. La cerca des de la categoria ‘Polítics’ (tasca 3), encara no l'hem triat dins la nostra mètrica d'anàlisi de categories (que són infermeria, biblioteconomia i educació), és una professió on es troben menys biografies de dones que homes a Wikipedia, i per això el cercador de gènere hauria de ser-li útil al participant del test a la tasca 6.

En el cas d'infermeres, a part de ser una professió feminitzada, es va considerar com una oportunitat per avaluar la intuitivitat de les etiquetes per a la persona usuària i si s'entenia bé la seva subclassificació. Hem de tenir en compte que “infermers” és categoria de primer nivell i “llevadores” també i pot generar una classificació complicada d'entendre. És el mateix cas que el d'escriptors (tasques 7-9), on la persona usuària s'enfrontava a una de les categories amb més subclassificacions.

Com que el nostre estudi té un enfocament no binària en termes d'identitat de gènere, testegem també cerques més específiques al cercador avançat, mitjançant l'arbre de categories, o el llistat jeràrquic d'una classificació concreta.

Taula 7. Experiments per al test amb les persones usuàries sobre les categories de Viquipèdia i les ontologies de Wikidata

Experiments	Descripció	Necessitat d'informació	Entorn	Mostra
-------------	------------	-------------------------	--------	--------

Tasca 1	Navegació autònoma	Dones del segle XX de l'àmbit educatiu	Viquipèdia	Persones usuàries: lectores i editores
Tasca 2	Cerca específica des del cercador general	Dona química catalana	Viquipèdia	Lectores i editores
Tasca 3	Cerca específica des del cercador de gènere a la categoria "polítics"	Dones a la política catalana del segle XX	Viquipèdia	Lectores i editores
Tasca 4	Cerca específica des de la cerca interna o avançada	Periodista no-binari	Viquipèdia	Lectores i editores
Tasca 5	Cerca específica des del cercador d'intersecció (PetScan)	Dona professora de l'àmbit STEM	Viquipèdia	Lectores i editores
Tasca 6	Navegació autònoma per l'arbre i/o l'índex de categories	Infermera catalana contemporània	Viquipèdia	Lectores i editores
Tasca 7	Navegació des de l'índex de categories	Dones escriptores catalanes contemporànies	Viquipèdia	Lectores i editores
Tasca 8	Navegació des de l'arbre de categories	Escriptors no-binaris catalans contemporanis	Viquipèdia	Lectores i editores
Tasca 9	Cerca amb el generador de consultes (QueryBuilder)	Escriptors no-binaris espanyols	Wikidata	Lectores i editores
Tasca 10	Navegació amb la llista de propietats (ontologies)	Cerca d'identificadors relacionats amb el gènere	Wikidata	Només editores

2.2.1.4. Elements interactius analitzats

En concret els elements interactius del sistema de navegació i cerca que s'analitzen són:

Navegació autònoma per la Viquipèdia: sense indicacions específiques per part de la moderadora de l'exercici, simplement se li indica a la persona usuària la necessitat d'informació que ha de satisfer i aquesta pot navegar intuïtivament utilitzant els recursos que consideri més adients, des de la portada o home page. [Tasca 1] (Vegeu Imatge 1)

Imatge 1. Portada de la Viquipèdia



Cerca específica a través del buscador general de Viquipèdia. [Tasca 2] (Vegeu Imatge 2)

Imatge 2. Cercador general



Cercador de gènere: cercador específic que només es pot trobar dins d'una categoria de professió o que designen a persones, com "polítics". Funciona recuperant la informació a partir de l'ontologia de Wikidata. És el cercador implementat per la comunitat de viquipedistes per solucionar la presa de decisió de no utilitzar les categories de "dones" i "persones no-binàries". [Tasca 3] (Vegeu Imatge 3).

Imatge 3. Cercador de gènere



Cercador intern: cercador amb opcions més avançades, on l'usuari/ària pot seleccionar més paraules i categories per trobar la cerca desitjada. [Tasca 4]. (Vegeu Imatge 4)

Imatge 4. Cercador intern

Resultats de la cerca

Q Cerca

Cerca avançada: ▼

— Cerca dins del text de la pàgina

Aquestes paraules ⓘ

Exactament aquest text ⓘ "camp de blat"

Cap d'aquestes paraules ⓘ

Una d'aquestes paraules ⓘ

— Estructura

El títol de la pàgina conté ⓘ

Subpàgines d'aquesta pàgina ⓘ

Pàgines dins d'aquestes categories ⓘ

Pàgines amb aquestes plantilles ⓘ

— Fitxers i imatges

Tipus de fitxer ⓘ Trieu el tipus de fitxer ▼

— Ordenació

Ordenació ⓘ Rellevància ▼

Cerca a: (Principal) × ▼

Cercador d'intersecció o interseccional: aquest és un cercador de Viquipèdia que va directe a fer consultes a Wikidata, entra d'altres bases de dades. És necessari buscar des de categories concretes, que són sensibles a l'accentuació i a l'ortografia. [Tasca 5]. (Vegeu Imatge 5).

Imatge 5. Cercador interseccional

Categories Propietats de la pàgina Plantilles i enllaços Altres fonts: Wikidata Resultat

Llengua: ca Wikimedia Commons Viquiespècies Wikidata

Projecte: wikipedia

Profunditat: 12 (profunditat 0 significa que no hi ha subcategories: si afegeix 1 i un número a la categoria en el camp següent, ignorarà i saltarà aquella categoria)

Categories: Polítics

Combinació: Subconjunt Unió

Categories negatives:

Executa [Link to a pre-filled form for the query you just ran with and without auto-run. PSID és 24808347. Cerca imatges per a Wikidata.](#)

resultats 6221

#	Títol	Identitat de la pàgina	Espai de noms	Mida (bytes)	Darrer canvi
1	Joana I de Castella	1579	(Article)	26986	20230408084859

Arbre de categories: és la visualització jeràrquica de l'organització d'una categoria amb relació a les altres per tal de veure les subordinacions i dependències. [Tasca 6 i 8] (Vegeu Imatge 6)

Imatge 6. Arbre de categories de Viquipèdia

Mostra l'arbre de categories

Categoria: Infermeria *

Mode: mostra només categories

Espai de noms: tots

Mostra l'arbre

Categories superiors: [Assistència sanitària](#) | [Categories amb enllaç commonscat des de Wikidata](#) | [Ciències de la salut](#) | [Professions sanitàries](#)

- ▼ Infermeria (3 cat., 18 p.)
 - ▶ Infermers (2 cat., 6 p.)
 - ▶ Llevadores (1 cat., 25 p.)
 - ▶ Primers auxilis (1 cat., 17 p.)

L'índex de categories: es tracta de la pàgina que recull les subcategories dins d'una categoria concreta i indica la quantitat d'articles i categories dins de cadascuna segons un índex alfabètic. [Tasca 7] (Vegeu Imatge 7)

Imatge 7. Índex de categories

D

▶ [Dramaturgs en català](#) (3 cat., 1 p.)

N

▶ [Novel·listes en català](#) (5 cat., 3 p.)

P

▶ [Poetes en català](#) (8 cat., 6 p.)

▶ [Presidents de l'Associació d'Escriptors en Llengua Catalana](#) (1 cat., 9 p.)

Consultes de Wikidata: **cercador específic a Wikidata.** Aquest cercador permet aconseguir resultats més concrets d'entre els continguts de Viquipèdia. La base de dades funciona amb propietats dins de les quals es troben altres propietats d'instància que s'engloben dins d'altra. [Tasca 9] (Vegeu Imatge 8)

Imatge 8. Cercador de Wikidata amb QueryBuilder

Consulta
Trobar tots els elements...

Propietat ⓘ
Introduïu una propietat

Valor ⓘ
Introduïu un valor

Referències ⓘ
amb i sense referències

Afegeix una condició

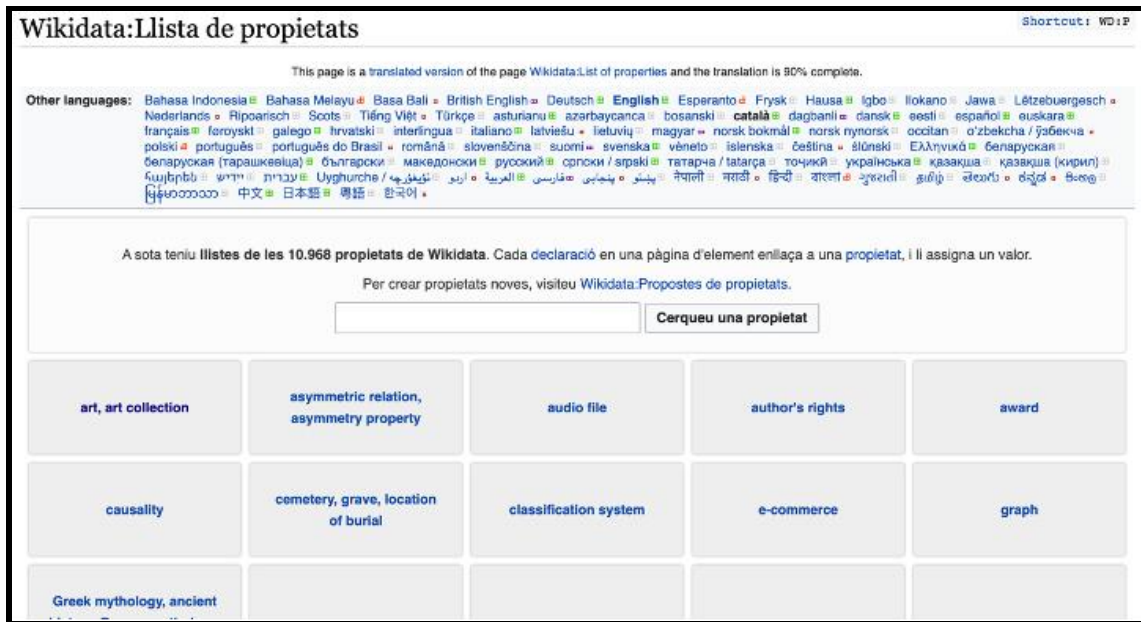
Paràmetres

Limit del número de resultats a 100

Mostra els identificadors en lloc de les etiquetes (pot alleujar el temps d'espera)

El llistat de propietats de Wikidata: permet la navegació per localitzar propietats través del vocabulari ontològic de Wikidata. [Tasca 10] (Vegeu Imatge 9)

Imatge 9. Llistat de propietats de Wikidata



2.2.1.5. Variables d'anàlisi

En aquesta secció es concreten les diferents variables d'anàlisi dels tests amb persones usuàries realitzats en la cerca i navegació de Viquipèdia. D'una banda, s'analitzen els aspectes relacionats amb la usabilitat i de l'altra amb l'experiència d'usuari.

La usabilitat d'un sistema fa referència a la facilitat amb què les persones poden usar una eina o un giny per aconseguir un objectiu concret. És a dir que valora el rendiment òptim de la interacció, la facilitat d'ús de l'eina sent intuïtiva, eficient i satisfactòria. Si una pàgina web és difícil d'utilitzar, la gent la deixarà de visitar. Per tant, és molt important assegurar que els projectes digitals tinguin una bona usabilitat.

Més específicament les variables que hem analitzat amb relació a la usabilitat i, en concret amb les qualitats relacionades amb l'ús són:

- Eficiència
- Eficàcia
- Facilitat d'aprenentatge
- Trobabilitat

I pel que fa a l'experiència d'usuari (UX) es té en compte tot un ventall d'emocions i d'experiències viscudes que s'engloben amb la variable de:

- Satisfacció

Finalment, seguint l'esquema metodològic de qualitat dels entorns webs de Morales-Vargas et. al. (2023), hem seleccionat les següents variables a partir dels autors Morville and Rosenfeld (2006) i Sauro and Lewis (2016):

Taula 8. Indicadors utilitzats per l'avaluació dels sistemes de navegació i cerca de Viquipèdia i Wikidata

Indicador	Referència	Descripció	Metodologia	Valor
Eficiència	Sauro and Lewis (2016)	Poden els usuaris aconseguir el seu objectiu amb el lloc web? La web dona el suport necessari per a què els usuaris puguin fer les seves tasques	Quantitatiu	Índex d'eficiència de navegació a partir del recorregut cognitiu i dels comportaments de les persones usuàries Probabilitat de retorn
Efectivitat	Morville and Rosenfeld (2006)	La web fa el que se suposa que ha de fer? El rendiment òptim de la interacció.	Qualitatiu Quantitatiu	Comentaris dels usuaris a partir del <i>think aloud</i>
Facilitat d'aprenentatge	Sauro and Lewis (2016)	Com de fàcil és la web d'aprendre a utilitzar-se? Que els usuaris triguin el menys possible a aprendre a utilitzar de manera eficient i sense cometre errors.	Qualitatiu	Valoració dels usuaris de l'experiència i intuïtivitat, en una escala del 0-5 Corba d'aprenentatge a partir dels clics fets al llarg del test per part de cada usuari.
Trobabilitat	Morville and Rosenfeld (2006)	Els serveis i productes han de ser localitzables, de manera que els usuaris puguin trobar el que necessiten.	Quantitatiu	Taxa d'èxit per tasca, càlcul del % de persones que satisfan la necessitat d'informació plantejada a la tasca
Satisfacció	Sauro and Lewis (2016)	Que puguin complir el seu objectiu amb el mínim esforç.	Qualitatiu	Valoració/satisfacció, en una escala del 0-5, comentaris del <i>think aloud</i> i respostes a les preguntes obertes

2.4. Inspecció d'estàndards i guies per a l'avaluació de la taxonomia (Viquipèdia) i les ontologies (Wikidata)

Com que no existeix actualment una metodologia referent que pugui determinar els indicadors òptims per l'avaluació d'una taxonomia o ontologia, sinó que aquestes mètriques es creen a partir de la rellevància de l'estudi pertinent, hem fet una revisió bibliogràfica dels estàndards més utilitzats per analitzar els sistemes d'organització del coneixement (SOC), conegudes amb les seves sigles en anglès, *Knowledge Organization Systems* (KOS), amb la qual poder dissenyar una guia d'avaluació dels criteris taxonòmics i ontològics per aplicar a l'estudi de Viquipèdia i Wikidata. I així aplicar el mètode d'inspecció d'estàndards per examinar si el SOC de Viquipèdia i

Wikidata segueix les especificacions i els estàndards referents definits per organitzacions internacionals.

Primerament, fent un estudi sobre el marc teòric del nostre estudi, seguim la proposta taxonòmica de Souza et al. (2012) per classificar els sistemes d'organització del coneixement utilitzant diferents dimensions. Així mateix, Mazzocchi (2018) ofereix una visió crítica introductòria sobre els sistemes d'organització del coneixement. Ambdós autors ens expliquen la lògica de classificació emprada en l'organització del coneixement, que hem afegit de referència per a la creació de la proposta d'avaluació de Viquipèdia i Wikidata.

En el cas concret de Viquipèdia, Albuquerque (2017) proposa un marc d'arquitectura de la informació per al cicle de vida de projectes de vocabulari de programació, per al desenvolupament i gestió de vocabularis controlats en aquest context específic. Més específicament, Kaplan et al. (2022) introdueixen un mètode d'avaluació per a taxonomies, on exposen una eina per mesurar i avaluar la qualitat de les taxonomies, ajudant a millorar la seva eficàcia i utilitat, que són els tres criteris de qualitat estructural: generalitat, apropiació-ateniment i ortogonalitat. Dins d'aquests quatre paràmetres, per avaluar i quantificar la generalitat i la idoneïtat d'una taxonomia, s'ha de servir de quatre mètriques generalitzades, com són la laconicitat, la lucidesa, la completesa i la solidesa. Finalment, Kless and Milton (2010) tracten la temàtica de les mesures de qualitat per avaluar tesaures, amb indicadors més concrets, que aporten una visió més àmplia que ens interessa al nostre estudi. Tots aquests els hem exposat de manera conjunta a l'Annex C.

En el context de les ontologies, da Costa et al. (2022) presenten una revisió actualitzada sobre les arquitectures de programari per a grans volums de dades, incloent-hi l'ús d'ontologies. A més, en el mateix context, Wilson et al. (2022) descriuen una metodologia per avaluar la qualitat d'una ontologia, tenint en compte els aspectes intrínsecs i extrínsecs. Amith et al. (2018) aborden l'avaluació d'ontologies, a l'àmbit específic de KOS biomèdiques, però que els criteris ens serveix per a l'avaluació de Wikidata. Bolotnikova et al. (2011) ens proposa uns mètodes pràctics per a l'avaluació de les ontologies, especialment en el context de l'automatització. Finalment, Aghaebrahimian et al. (2022) exploren la validesa de les categories de Wikipedia per l'etiquetatge de temes utilitzant categories de Wikipedia. Unificant les teories d'aquestes, hem establert la taula fixada a l'Annex D.

Ambdues taules són una aproximació als criteris que hem considerat per abastar les dimensions més rellevants a l'estudi de la categorització amb perspectiva de gènere, resoldre el dubte de si el problema que pateix la Viquipèdia és estructuralment mesurable i si Wikidata és la resposta més adequada per a resoldre-ho. L'avaluació que hem finalment determinat està desglossada al següent apartat.

2.5. Proposta d'avaluació heurística de l'esquema de categories de Viquipèdia

L'avaluació heurística, amb el fi d'analitzar si la taxonomia de Viquipèdia respectivament compleix amb els estàndards d'una bona organització del coneixement, no només respecte a l'experiència de l'usuari, sinó formalment del KOS. Basant-nos en el marc teòric prèviament explicat organitzat en les taules imposades a l'Annex C, hem seleccionat els indicadors, per una part, més destacats per la nostra anàlisi i, per l'altra, que siguin aconseguibles amb la nostra capacitat d'accés i equipament tècnic disponible (Vegeu Taula 9). Com a fonts per a la identificació i mesura d'indicadors hem tingut en compte les aportacions dels especialistes i dels estàndards específics del sector de SOC i, específicament, taxonomies per tal de procedir a fer una anàlisi d'inspecció de l'esquema de categories de Viquipèdia (Vegeu Taula 9).

Taula 9. Indicadors utilitzats en el mètode d'inspecció amb estàndards a l'esquema de categories de Viquipèdia

Indicador	Referència	Descripció	Metodologia	Valor
Avaluabilitat	Alós-Moner et al. (2010)	Existeixen mecanismes d'avaluació que permeten determinar els nivells de qualitat de l'esquema de categories, i detectar desviacions al llarg del temps.	Existència de procediments consensuats, aprovats i difosos. Confrontació dels procediments amb el contingut de l'esquema de categories.	Valor binari
Reusabilitat	Alós-Moner et al. (2010) GGPT (2009)	L'esquema de categories ha de ser útil en diferents escenaris de classificació i ús dins de la Wikipedia, presa en conjunt o de manera parcial. El grau de reutilització en cada context dependrà, en gran mesura, de les exigències d'especificitat i exhaustivitat d'aquest context. Quin format d'intercanvi de dades està disponible per a l'extracció i implementació de l'esquema de categories.	Existència de procediments consensuats, aprovats i difosos. Confrontació dels procediments amb el contingut de l'esquema de categories.	Valor binari
Estabilitat	Alós-Moner et al. (2010)	L'estructura i els conceptes escollits han de ser perdurables, tret que els requeriments d'actualització contínua recomanin la incorporació de canvis. En cap cas no s'inclouran categories que requereixin actualització temporal	Anàlisi de dades temporals sobre creació de categories.	Valor binari

		(per exemple, pressupost vigent).		
Nombre de categories (cocneptes)	Alós-Moner et al. (2010) Stock (2015)	Còmput de les categories (conceptes) en el SOC, i comparació amb recursos similars. La mitjana de documents classificats per categoria és un indicador complementari d'aquesta dimensió.	El còmput es realitza a partir del bolcat de les dades relatives a les categories de Wikipedia (<i>database dump</i>). Es realitza la comparació amb llistes d'encapçalaments de biblioteques i amb enciclopèdies originalment publicades en suport analògic.	Comparació
Nombre de relacions semàntiques	Alós-Moner et al. (2010) Stock (2015)	Còmput de les relacions semàntiques entre categories (conceptes) en el SOC.	El còmput es realitza a partir del bolcat de les dades relatives a les categories de Wikipedia (<i>database dump</i>)	Cas d'estudi a partir del bolcat de base de dades
Índex d'enriquiment o granularitat	Alós-Moner et al. (2010) Gil Leiva (2008) Lancaster (2002) Stock (2015)		Mitjana entre el nombre total de relacions i el nombre de categories. Les referències situen el nombre màxim de nivells entre 2 i 5.	Valors òptims
Grau de precoordiació	Alós-Moner et al. (2010) Lancaster (2002) Stock (2015)	La preccordinació consisteix en la Combinació de conceptes en el moment de la creació de categories o en el momento d'utilitzar-les en la categorització. S'oposa a la postcoordinació que consisteix en la combinació dels conceptes, per part dels usuaris, en el moment de la cerca.	El còmput es realitza a partir del bolcat de les dades relatives a les categories de Wikipedia (<i>database dump</i>). Es calcula la mitjana entre el nombre de paraules significatives (noms, adjectius i verbs) de les categories i el nombre total de categories. Les referències situen el nombre màxim de nivells entre 1,5 i 2.	Cas d'estudi a partir del bolcat de base de dades

Nombre de nivells en la jerarquia o profunditat	Alós-Moner et al. (2010) Stock (2015)	Es tenen en compte les categories vinculades per la relació de jerarquia, en una mateixa cadena, des del nivell superior fins al nivell més baix.	Mitjana entre el nombre total de nivells i el nombre de categories. Les referències situen el nombre màxim de nivells en 5.	Valors òptims
Nombre de categories en un mateix nivell de la jerarquia o amplitud	Alós-Moner et al. (2010) GGPT (2009) Stock (2015)	Es tenen en compte les subcategories de totes les categories, des del nivell superior fins al nivell immediatament anterior al més baix.	Mitjana entre el sumatori de subcategories, i el nombre total de categories (excepte les categories de darrer nivell). Les referències situen el nombre mínim en 2, i el nombre màxim 12.	Valors òptims

2.6. Anàlisi dels processos de presa de decisions

Les categories a la Wikipedia són creades per les persones editores, i mantingudes o esborrades pels administradors¹⁹, que són persones electes per la comunitat a través de votació i un llindar considerable de producció de continguts. El sistema de categorització per *folksonomies* és creat per aquest col·lectiu de forma descentralitzada (Vander Wal, 2007), ja que és el seu rol, però no impedeix que una persona creï una categoria (per exemple “persones LGBT” (Categoria, s. f.) i “persones trans”), encara que després s'esborri de forma immediata. I en algunes ocasions es pot proposar un procés de deliberació i fins i tot executar el procediment de votació previst per Wikipedia per decidir la inclusió o el rebuig.

En el cas de les categories “dones” i “persones no-binàries” de la Viquipèdia, la presa de decisions es va produir amb tres processos de deliberació i en un procediment de vot²⁰ on es va resoldre no acceptar les categories que distingeixen per gènere. En concret de l'etiqueta de classificació “dones” va rebre dues deliberacions una l'any 2014, a partir de la creació de la categoria “dones” per part d'un administrador, i una altra entre el 2015 i el 2017, realitzades totes dues a La Taverna, espai de discussions de la Viquipèdia. Posteriorment, el 2018 es va proposar una votació per decidir l'adaptació de categoria de gèneres a l'enciclopèdia online. Mentre que la primera deliberació va durar 27 dies i la segona entre el 2015-17, amb un debat obert durant gairebé dos anys, la deliberació al voltant de la categoria “persones no-binàries” es va debatre durant un sol dia.

Així mateix, després de la creació autònoma per a una persona usuària de la categoria “persones no-binàries”, es va dur a terme una altra deliberació arran de la seva eliminació el 2022, debat que trobem a la pàgina de la persona usuària en qüestió. Per tant, el corpus utilitzat en l'anàlisi del contingut prové dels textos de les deliberacions de la Taverna, més una pàgina d'usuari:

- Categoria “Dones”:
 - Taverna, Categorització segregada per gèneres, altre cop (2015)²¹

¹⁹ https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Presade_decisions

²⁰ https://ca.wikipedia.org/wiki/Categoria:Converses_sobre_la_categoritzaci%C3%B3_de_les_dones

²¹ <https://ca.wikipedia.org/wiki/Tema:Su3sim7j2skj1883>

- Taverna, Categoria Dones (2014)²²
- Categoria “Persones no-binàries”
 - Pàgina Usuari Brunnaiz (2022)²³
- Categories per identitats de gènere:
 - Votació (2022)²⁴

Aquest corpus se n'ha recollit, analitzant-ne el contingut a partir del llibre de codis de tipologia d'arguments proposada per Schneider et al. (2013) sense ser necessari cap adaptació especial a les categories i definicions. Simplement, s'ha incorporat un nou element d'anàlisi que incorpora la perspectiva de gènere. El llibre de codis utilitzat conté les categories següents per a cada tipologia d'arguments (Taula 10).

Taula 10. Llibre de codis d'anàlisi de les deliberacions de les categories d'identitats de gènere “dones” i “persones no binàries” a la Viquipèdia

	Categoria	Descripció
1	Analogia (<i>Argument from Analogy</i>)	Quan es fa referència a casos similars. Per exemple, es compara amb una biografia d'una persona homòloga o es compara amb altres decisions preses a la Viquipèdia prèviament.
2	Valors (<i>Argument from Values</i>)	S'ofereix un judici de valor a l'aportació. Per exemple, es donen valoracions amb valors propis o se'n demerita o exalta el valor amb criteris personals.
3	Sense raonament (<i>Not reason given</i>)	No es donen explicacions pròpies, només es fa una votació a partir de comentaris oferts anteriorment.
4	Causa-efecte (<i>Cause-Effect</i>)	Quan falten fonts d'informació, manca estructurar l'article, la redacció no és prou enciclopèdica, etc. aleshores l'efecte és la decisió exposada.
5	Normes o regles de Wikipedia (<i>Arguments from Rules</i>)	Quan es fa referència a les polítiques i lleis de Wikipedia com poden ser la política de rellevància (WP:N) o al punt de vista neutral (WP: NPOV).
6	Raonament pràctic (<i>Practical reasoning</i>)	Es duen a terme accions cap a una meta o objectiu, és a dir, solucions pràctiques o més fàcils d'implementar per aconseguir consens o la fi del debat.
7	Desconeixement o ignorància (<i>Argument from unknowing/ignorance</i>)	Es fan suposicions quan no es pot trobar evidència o expressen el desconeixement sobre el tema o el camp del coneixement sobre què es debat.
8	Classificació verbal (<i>Argument from verbal classification</i>)	S'utilitzen arguments que usen o citen definicions per a conceptes concrets.

²² https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:La_taverna/Arxius/2014/Febrer

²³ https://ca.wikipedia.org/wiki/Usuari_Discussi%C3%B3:Brunnaiz/Arxiu_1

²⁴

https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Preses_de_decisions/2018/Sobre_la_categoritzaci%C3%B3_per_g%C3%A8ner

9	Rebuig (<i>Argument from waste</i>)	Quan es pretén evitar la pèrdua de feina o esforç. Ho incloc quan fan referència al fet que és un malbaratament perdre l'esforç invertit.
10	Necessitat d'ajuda (<i>Argument from need to help</i>)	Se sol·licita ajut a altres persones editores.
11	Biaix (<i>Argument from bias</i>)	Quan es considera que és promocional o hi ha un conflicte d'interès.
12	Evidència cap a la hipòtesi (<i>Argument from evidence to hypothesis</i>)	Es proveeix evidència, inclou dades contrastades com a exemple o evidència del que s'ha argumentat o s'adjunten enllaços o referències.
13	Posició de coneixement (<i>Argument from position to know</i>)	Qui edita coneix professionalment el tema que es debat.
14	Precedents (<i>Argument from precedent</i>)	Es fan referències a discussions passades i es genera una "jurisprudència" de Wikipedia que guia sobre com actuar en la mateixa situació.
15	Gènere	Els comentaris proposats a la deliberació es fonamenten explícitament en raons de discriminació de gènere. Categoria afegida al llibre de codis de Schneider et al. (2013).

A les deliberacions, el gènere de les persones editores es va codificar també seguint la metodologia anteriorment utilitzada a Minguillón et al. (2021). El procés de codificació es va realitzar entre maig i juny del 2022 mitjançant el programari Atlas.ti (Hwang, 2008). L'anàlisi de fiabilitat (reliability) mitjançant el coeficient Kappa de Cohen (1960) va mostrar un nivell d'acord del 92,4%. Finalment, s'han anonimitzat els noms de les persones usuàries van participar en les votacions i deliberacions.

L'anàlisi de la presa de decisions a Wikidata s'ha fet a partir d'una revisió de la literatura, ja que ja hi ha un projecte de recerca investigant aquest aspecte (WIGEDI)²⁵.

3. Revisió de la literatura

En aquesta secció, oferim una revisió exhaustiva de la literatura seguint el marc SALSA (Grant i Booth, 2009) per examinar la bretxa de gènere a la Viquipèdia i a Wikidata. La recerca acadèmica ha investigat extensament la bretxa de gènere en ambdues plataformes. La fase d'avaluació de la Viquipèdia va involucrar 97 articles, i l'avaluació de Wikidata va involucrar 34. Un total de 21 articles van ser utilitzats per avaluar la Viquipèdia (Ferran-Ferrer et al., 2023), i 19 van ser utilitzats per avaluar Wikidata.

²⁵ <https://wigedi.com/>

3.1. Wikipedia

La bretxa de gènere a Wikipedia ha estat objecte d'àmplies investigacions acadèmiques, amb nombrosos estudis que exploren els prejudicis en el contingut, la participació, la lectura i les estratègies potencials per abordar aquesta bretxa. Aquests estudis posen èmfasi en la importància de reconèixer i abordar els prejudicis i les barreres per crear una comunitat de Wikipedia més diversa i inclusiva.

La subrepresentació de les dones com a editores i com a temes de cobertura biogràfica és un problema àmpliament reconegut en l'àmbit acadèmic. Alguns articles discuteixen com el prejudici de gènere s'entrellaça amb la raça, la sexualitat, la seguretat i la marginació a Wikipedia. Diversos factors, com la demografia de les persones editores, l'estructura de la plataforma i els valors culturals contribueixen a aquests prejudicis, que tenen implicacions significatives socialment, afectant la visibilitat i la participació de les dones i perpetuant les disparitats existents.

Pel que fa a la bretxa de gènere en el contingut, les investigacions revelen que les dones estan subrepresentades entre les principals figures de totes les edicions en llengua de Wikipedia (Miquel-Ribe & Laniado, 2021). Les biografies de dones i de persones LGBTQ+ sovint són objecte de supressió, la qual cosa provoca una major proporció de biografies de dones nominades per a la supressió en comparació amb les biografies disponibles sobre dones (Morgan et al., 2013; Hollink et al., 2018; Tripodi, 2023). No obstant això, el contingut que poden ser d'interès per als homes sembla que tenen més probabilitats de ser nominat per a supressió ràpida. Tot i que hi ha indicis de prejudicis, alguns autors conclouen que no és evident cap prejudici que resulti de l'activitat de supressió (Worku et al., 2020).

Els estudis també identifiquen diferències significatives de gènere en el contingut de Wikipedia, com ara el fet que les biografies de dones presenten més prominència en temes familiars, de gènere i de relacions (Wagner et al., 2016). Es pot observar un prejudici lingüístic en termes d'abstracció i positivitat del llenguatge, i hi ha diferències estructurals en les metadades i els enllaços hipertextuals. A més, les pràctiques de citació revelen que les dones autors són citades menys del que s'esperaria, el que suggereix una preferència per citar publicacions d'homes (Zheng et al., 2022). Aquests prejudicis poden minoritzar encara més les dones autores, sobretot en països no anglosaxons. La bretxa de gènere en la creació de contingut i la participació a Wikipedia perpetua una cobertura desequilibrada dels temes, creant un cercle en què la manca de diversitat en el contingut no aconsegueix atraure i implicar diferents persones editores, exacerbant així la bretxa de gènere existent.

La investigació sobre la bretxa de gènere en l'edició i la participació destaca diverses barreres que dificulten la implicació de les dones a Wikipedia. Aquestes barreres inclouen la reputació negativa, la falta de reconeixement, la por d'eliminació, el rebuig i l'alienació. Sovint, les recerques plantegen que les dones no de confiança en les seves habilitats, se senten incòmodes amb l'edició i enfronten respostes negatives a la crítica constructiva (Collier & Bear, 2012). Factors com la bretxa d'habilitats informàtiques (Gardner, 2011), i la disponibilitat de temps per a l'edició (Gruwell, 2015) també contribueixen a la bretxa de gènere. No obstant això, les editores femenines visibles i els comentaris constructius poden ajudar a mitigar la bretxa, ja que la presència de companyes femenines visibles fomenta l'edició col·laborativa (Evans et al., 2015). Alguns autors, com Buchem and Kloppenburg (2013), han investigat la bretxa de gènere a Alemanya i han suggerit un enfocament proactiu per a la formació i l'ensenyament de les dones, amb l'objectiu de potenciar la seva motivació per escriure. Han destacat l'impacte de les responsabilitats familiars en la capacitat de les dones per escriure, per la qual cosa potser els esforços haurien de centrar-se en abordar les disparitats de gènere en la feina domèstica (Ferran-Ferrer et al., 2021).

La bretxa de gènere s'estén més enllà de l'edició i inclou la manca de representació de persones amb orientacions sexuals no heterosexuales (Miquel-Ribé & Laniado, 2021). Accedir a citacions fiables que compleixin les polítiques de Wikipedia presenta reptes en el procés d'edició.

Les administradores femenines tenen un paper significatiu en fomentar un ambient inclusiu a Wikipedia, i hi ha diferències notables en el seu enfocament en comparació amb els administradors masculins (Farzan et al., 2016; Karczewska & Kukowska, 2021). La participació femenina varia segons el tema, amb una major presència en les categories d'estudis de gènere o feminisme, reflectint estereotips de gènere tradicionals. Les restriccions genèriques del lloc limiten la credibilitat digital i l'autoritat de les dones, dificultant les seves contribucions. La relació complexa entre la bretxa de gènere i l'assetjament requereix una millor comprensió, i és important crear un entorn segur per a les dones a dins i fora de Wikipedia. Les intervencions feministes, com els edit-a-thons exclusius per a dones, han demostrat ser efectives en contrarestar la desigualtat de gènere a la plataforma.

Els esforços per abordar la bretxa de gènere i la marginació a Wikipedia impliquen diverses intervencions i consideracions. L'assessorament femení s'identifica com un factor crucial per promoure la inclusió de les dones. No obstant això, també és important abordar la disparitat de gènere entre les persones editores influents que configuren les polítiques i duen a terme tasques de nivell superior. Tot i que s'ha progressat en l'atracció de més dones a Wikipedia, és necessari assolir una major paritat de gènere entre les persones editores influents (Antin et al., 2011). La bretxa de gènere a Wikipedia té conseqüències àmplies i importants que van més enllà del món digital. Reforça els estereotips de gènere existents, perpetuant la idea que les contribucions dels homes són més valuoses i notables que les de les dones. Aquesta representació tendenciosa no només distorsiona la nostra comprensió de la història, sinó que també obstaculitza el progrés cap a la igualtat de gènere. En esborrar les contribucions de les dones dels registres històrics, reforcem la idea que les dones no han jugat un paper important en la configuració de les nostres societats, perpetuant així la desigualtat social en el procés.

La bretxa de gènere a Wikipedia és un tema de preocupació per les seves conseqüències transcendents més enllà de la representació en línia. Aquesta bretxa no només perpetua els estereotips de gènere, sinó que també esborra les contribucions de les dones dels registres històrics i reforça la desigualtat social (Acey et al., 2021). Els algoritmes tendenciosos exacerbant encara més el problema fent difícil trobar informació sobre les dones, infravalorant els seus assoliments en el procés (Wagner et al., 2015). Per abordar eficaçment aquesta bretxa de gènere, és imprescindible dedicar recursos suficients per incloure perspectives diverses i assegurar una representació justa del coneixement a la plataforma. D'aquesta manera, podem lluitar activament contra la discriminació i promoure una producció de coneixement més equitativa a Wikipedia.

Per abordar aquesta bretxa de gènere, és crucial dedicar recursos suficients per incloure perspectives diverses i garantir una representació equitativa a la plataforma. Això implica promoure activament la participació de les dones com a editores i impulsar la inclusió de contingut relacionat amb les dones. També cal examinar críticament els algoritmes i les pràctiques de classificació per garantir que no perpetuïn els prejudicis de gènere. És necessari un esforç col·lectiu per abordar aquesta bretxa i treballar cap a una producció de coneixement més equitativa i representativa a Wikipedia.

Per una revisió més aprofundida es pot consultar Ferran-Ferrer, *et al.* (2023).

3.2. Wikidata

La revisió sistemàtica de la literatura a través de l'*scoping review* no ha facilitat la descoberta de gaires estudis sobre gènere i Wikidata, de fet la base de dades apareix a finals del 2012²⁶ i els primers autors a aproximar-se al tema van ser Klein et al. (2016) i Konieczny and Klein (2018). Aquests darrers van apuntar a la dificultat de mesurar la representació o l'absència de les biografies de dones a l'enciclopèdia i moguts per aquesta preocupació van generar el "Wikidata Human Gender Indicators" (WHGI) basat en dades de Wikidata, que aporta un conjunt de dades longitudinal, biogràfic, al llarg del temps i l'espai, cultures, llengües i professions, entre 11 elements en total d'estudi (Konieczny & Klein, 2018; Zhang & Terveen, 2021). Aquesta recerca confirma, amb dades, que la bretxa de gènere és un fenomen que s'ha produït al llarg de la història i de les cultures.

Diversos autors coincideixen en el fet que Wikidata aporta solucions a l'estudi de la bretxa de gènere a Wikipedia (Mora-Cantalops et al., 2019; Zhang & Terveen, 2021). Així, Zhang and Terveen van treballar per quantificar la bretxa de gènere i van poder especificar que les dones a Wikidata eren menys del 22%. Van voler estudiar si el biaix es produïa a causa de les persones voluntàries que editen Wikidata o bé si era per l'efecte mirall del *dataset* que reflecteix la realitat. Van poder concloure que Wikidata no incrementa la bretxa de gènere, a través d'un estudi que comparava dades de professions de col·legis professionals amb dades d'ocupacions de Wikidata. Entre les dades de l'estudi van aconseguir quantificar, per exemple, que entre les professions més populars entre els homes a Wikidata són els jugadors de futbol. És a dir que Wikidata reflecteix la realitat tal com és sense augmentar el biaix ja existent, si es compara amb Viquipèdia, l'oportunitat de reproduir biaixos en un article és més alta que l'oportunitat que suposa aportar dades concretes sobre una persona a Wikidata. Conjecturem que això no ofereix tantes oportunitats perquè es produeixin prejudicis de gènere, fins i tota mena de prejudicis implícits, com s'ha demostrat en comparacions múltiples de l'ús del llenguatge en biografies de la Viquipèdia. A més, gran part del contingut de Wikidata s'afegeix mitjançant bots automatitzats²⁷, que importen informació de fonts de dades externes com ara l'Encyclopaedia Britannica. Això també contribueix a aportar tipus i quantitats equivalents d'informació factual sobre dones i homes. L'estudi aporta també possibles explicacions d'aquestes troballes i possibles implicacions per reduir la bretxa de gènere a Wikidata (Zhang & Terveen, 2021).

Des de les recerques analitzades, la utilització de Wikidata es considera una oportunitat important doncs per augmentar la diversitat i mitigar la desigualtat en el contingut de la Wikipedia (Klein et al., 2016; Hollink et al., 2018). Wikidata, és un conjunt de dades lliures i col·laboratives que serveixen com a dipòsit central per a dades estructurades que es poden utilitzar per enriquir els articles de la Wikipedia amb una àmplia gamma d'informació. Mitjançant l'aprovació de les extenses dades de Wikidata sobre diversos temes, les persones editores poden millorar la diversitat i inclusió del contingut de la Wikipedia.

Reconeixent la necessitat de tenir una representació més inclusiva, algunes publicacions han intentat utilitzar fonts externes com ara Wikidata o DBPedia per abordar el buit de contingut en identitats de gènere (Wagner et al., 2016). Aquestes fonts proporcionen una gran quantitat de dades que es poden utilitzar per ampliar la cobertura de les identitats de gènere i assegurar que les persones de diferents orígens estiguin adequadament representades en els articles de la Wikipedia. Mitjançant la incorporació d'informació de fonts externes com ara Wikidata, les persones editores tenen accés a una gamma més ampla de perspectives i poden contribuir a una comprensió més completa del gènere i la identitat.

²⁶ <https://en.wikipedia.org/wiki/Wikidata>

²⁷ <https://wikidata.org/wiki/Wikidata:Bots>

No obstant això, malgrat el potencial que ofereix Wikidata, encara hi ha limitacions en la seva utilització per aconseguir una diversitat completa a la Wikipedia. Tot i que molts articles biogràfics de la Wikipedia tenen entrades corresponents a Wikidata, encara hi ha casos en què les persones no estan registrades com a éssers humans a Wikidata (Klein et al., 2016). Aquesta discrepància pot fer que determinades biografies siguin ignorades o excloses de la base de coneixement més ampla, dificultant els esforços per promoure la diversitat i la igualtat.

Per abordar aquests reptes, s'estan realitzant esforços per millorar la integració entre la Wikipedia i Wikidata. Iniciatives com ara Women in Red²⁸, Wikimujeres²⁹, Viquidones³⁰, etc. se centren a augmentar la representació de les dones a la Viquipèdia creant nous articles i ampliant els existents. Mitjançant l'aprofitament de la informació disponible a Wikidata, aquestes iniciatives tenen com a objectiu superar els prejudicis i garantir que les persones poc representades rebin el reconeixement que els correspon.

En conclusió, si bé la utilització de Wikidata té un gran potencial per augmentar la diversitat i mitigar la desigualtat en el contingut de la Viquipèdia, encara hi ha obstacles a superar. Mitjançant l'aprofitament de fonts externes com ara Wikidata, les persones editores poden ampliar la cobertura d'identitats diverses. No obstant això, el registre incomplet de persones a Wikidata planteja un repte per aconseguir una representació completa. La col·laboració i la millora continua entre la Viquipèdia i Wikidata poden conduir a una base de coneixement més inclusiva i diversa que reflecteixi la riquesa de les experiències humanes.

Generalment, la recerca en Wikidata s'ha centrat a analitzar la qualitat de les dades de Wikidata en termes de qualitat de les dades o avaluació de l'ontologia, però habitualment s'avalua sense tenir en compte la satisfacció de les necessitats de les persones usuàries que són qui acabaran utilitzat el recurs d'informació. Les persones autores de l'article proposen un marc d'avaluació de les ontologies tenint en compte les mancances fonamentals en els continguts (content gaps) i els compara en l'àmbit qualitatiu amb els desequilibris del knowledge graph amb les necessitats d'informació de les persones usuàries. L'estudi se centra en Wikidata, examinant específicament les mancances de gènere, actualitat, geogràfiques i socioeconòmiques. L'estudi conclou que les mancances de gènere i actualitat no són intrínseques a Wikidata, però hi ha ocasions en què les persones editores descuiden les entitats poc representades. Els autors esperen que aquest estudi ajudi els enginyers del coneixement a identificar i abordar les mancances de contingut (Abián et al., 2022).

Un altre estudi que d'alguna manera també se centra en les necessitats de les persones usuàries és el del marc d'anàlisi de dades que proporciona WikiTrends³¹ amb totes les visualitzacions que ofereix a partir de coordenades espacial/temporals, per professions, per gènere, etc. d'un conjunt de gairebé 5 milions de pàgines (Guerguis et al., 2017).

L'article sobre Pantheon 1.0 (Yu et al., 2016) explica un exercici comparatiu per analitzar la qualitat de les dades de Wikidata amb relació a mesurar la popularitat de més de 10 mil biografies de 25 llengües diferents.

Un dels articles analitzats aporta informació sobre un sistema que utilitza Wikibase com a interfície amigable per tal que els humans accedeixin a un model de dades enllaçades subjacent i que també es pot utilitzar com a servei de publicació de dades on aquests experts poden navegar fàcilment i afegir els seus coneixements. L'article proposa a més un sistema que sincronitzi automàticament els fitxers RDF allotjats en un sistema de control de versions amb una instància

²⁸ https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red

²⁹ <https://es.wikipedia.org/wiki/Wikiproyecto:Mujeres>

³⁰ <https://www.upf.edu/web/igualtat/viquidonesupf>

³¹ <https://www.wikitrend.org/>

de Wikibase. Es descriu el sistema des d'un punt de vista arquitectònic, i expliquen els principals components necessaris per a la sincronització de dades. També s'aborden alguns dels reptes per a una sincronització efectiva dels fitxers RDF amb Wikibase. Aquest sistema s'està utilitzant actualment com a part del projecte Hèrcules per sincronitzar una ontologia d'ecosistema de recerca amb una instància de Wikibase (Hevia et al., 2020).

Hi ha un estudi que planteja la identificació de classes de Wikidata a partir dels continguts de la Wikipedia en diversos idiomes (Pasca, 2018) per tal de facilitar la feina de generar dades a Wikidata.

Un cas d'estudi recent que ha destacat la importància de la col·laboració i el treball comunitari en projectes que enriqueixen les dades obertes vinculades, les dades semàntiques al voltant del virus i de la pandèmia. El projecte utilitza esquemes d'entitats, descripcions en llenguatge natural i formats basats en web per millorar l'accessibilitat de les dades de la COVID-19. Les contribucions dels participants han donat lloc a un recurs valuós per a la comunitat de ciències de la vida per accedir al coneixement acumulat i a la informació actualitzada sobre la pandèmia de COVID-19 ("wikidata WikiProject COVID-19")³². És un bon exemple per copsar el dinamisme de la comunitat editora de Wikidata i de l'impacte social que pot arribar a tenir.

En relació amb el tema que ens ocupa en aquest informe, la visualització de les diferents identitats de gènere, és molt interessant el treball de Weathington and Brubaker (2023), el qual explica com la recopilació, l'organització i la recuperació de dades sobre individus queer i les seves identitats desafien els creadors i curadors de sistemes de bases de dades altament estructurats. Tot i que les investigacions prèvies en estudis d'arxiu i demografia han examinat els processos de recollida i emmagatzematge d'identitats queer, no examinen les complexitats de plataformes altament democratitzades que no tenen mandats de dalt a baix que sovint estructurin esquemes d'arxiu. Per examinar la representació de persones queer en bases de dades de plataforma oberta, es va realitzar una etnografia de rastres i una anàlisi temàtica de Wikidata, una base de dades de col·laboració oberta i altament estructurada. Van examinar específicament la creació, els canvis, les discussions i els impactes de propietats que codifiquen identitats queer, com ara l'orientació sexual i el sexe o el gènere. Van trobar que els canvis sovint tenen impactes inesperats, que els col·laboradors lluitaven per determinar el vocabulari de les identitats queer que fos precís en els diversos contextos culturals de la comunitat de Wikidata, que l'enregistrament d'identitats queer podria causar un efecte estigmatitzador per a les persones LGBTQ+, amb més preocupacions de difondre rumors o sortir a la gent tancada, i que els col·laboradors proposin canvis que provocarien representacions esbiaixades de persones queer. L'anàlisi demostra friccions inherents i no abordades en traduir identitats queer als límits d'una base de dades estructurada. Conclouen discutint maneres en què la naturalesa col·laborativa altament de baix a dalt de plataformes com Wikidata, sovint vista com una fortalesa important, pot ser vulnerable a què individus o grups petits descarrilin i obstrueixin els canvis amb els quals no estan d'acord en temes amb càrrega política com ara les identitats queer.

Una recerca menys empírica que les anteriors, però igualment necessària, ressalta la importància d'estudiar els prejudicis i biaixos de les persones i com es traslladen a les tecnologies. No se centra exclusivament en Wikidata però explora els avantatges i els inconvenients dels algorismes i les noves tecnologies en sistemes d'organització dels continguts com pot ser Wikidata. Posa èmfasi en el fet que, tot i que aquestes tecnologies ajuden les persones en diferents situacions de la vida, també poden generar discriminació i prejudicis. Els autors presenten el mètode del "Feminisme de Dades (o Data Feminism)" com un enfocament per mitigar els prejudicis en els projectes de ciència de dades. L'article avalua la viabilitat d'implementar aquest mètode, posant èmfasi en els punts de vista subjectius, atorgant crèdit a

³² https://wikidata.org/wiki/Wikidata:WikiProject_COVID-19/en

totes les formes de treball i construint comunitats mitjançant esdeveniments de marcatge públic. El projecte utilitza fonts de dades obertes com ara Wikidata i OpenStreetMap. Els autors tenen com a objectiu fomentar que les persones considerin els prejudicis en les seves dades i adoptin enfocaments més ètics en els projectes de ciència de la computació (Gleißner et al., 2021).

4. Objectius

En algunes Viquipèdies com la italiana o la catalana, hi ha específicament un biaix pel que fa a l'accés i la visualització de continguts sobre la diversitat d'identitats de gènere. En aquests casos, categories com "dones" o "persona no binària" estan prohibides per a l'organització i la recuperació de la informació. Aquestes decisions desenvolupen algunes disfuncions que són especialment crítiques en llengües que utilitzen el gènere gramatical.

En aquest informe es pretén fer una contribució a la recerca reduint la bretxa de gènere a la Viquipèdia mitjançant la visualització de les diversitats de gènere recollides a Wikidata. Per tant, busquem proporcionar un motor de cerca i un sistema de navegació utilitzant les ontologies de Wikidata per tal de proporcionar accés i més visualització de la diversitat d'identitats de gènere a la Viquipèdia.

Així, mentre que a Viquipèdia només es pot optar per formes masculines en les categories, a Wikidata els continguts sobre gènere (propietat 21) recullen diferents classes relatives a gènere, com ara agènere, dona, home, intersexual, no-binari, transgènere dona i transgènere home, entre d'altres.

Per tant, la solució que proposem a la dificultat de la visualització de les diferents identitats de gènere i a la recuperació dels seus continguts, apunta cap a utilitzar les ontologies de Wikidata que podrien aportar organització i una millor representació del que es coneix a la Viquipèdia. De fet, ja hi ha precedents d'aquesta implementació en alguns casos, la Viquipèdia ja utilitza categories que sorgeixen de Wikidata, com ara "persones vives"³³.

Wikidata té el repte particular de modelar el gènere com a dades estructurades. Aquesta proposta de projecte pretén presentar un cas concret (a la Viquipèdia en català) de la bretxa de gènere en continguts (biografies) i categories de Viquipèdia per a una millor visualització i accés a la diversitat d'identitats de gènere a través de les ontologies de Wikidata.

Tanmateix, Wikidata no és més esbiaixada que el món real (Zhang and Terveen, 2021). És per això que el plantejament del problema i les possibles solucions segueix una doble aproximació, d'una banda, tecnològica i de l'altra cultural. La tecnologia no és neutral i ambdues aproximacions s'influeixen mútuament. La solució ha de venir de tots dos.

La proposta de projecte produirà un diagnòstic mitjançant l'anàlisi del sistema d'organització del coneixement de la Viquipèdia (categories/folksonomies) i Viquidata (ontologies) des d'una perspectiva de gènere. L'enfocament serà tecnològic però també cultural en els processos de presa de decisions. A continuació, es recolliran els requisits tècnics dels sistemes de cerca i organització del coneixement (Viquipèdia, WikiBase) i els requisits dels usuaris. I per últim, però no menys important, la proposta inclou el desenvolupament d'un prototip per a provar a la Viquipèdia en català.

³³ https://en.wikipedia.org/wiki/Category:Living_people

Taula 11. Objectiu general de la proposta

Objectiu general
L' objectiu general d'aquest informe és el d'analitzar el biaix de gènere a les categories de Viquipèdia a través de realitzar una diagnosi tècnica, posicionament acadèmic, per poder finalment dissenyar una proposta de millora sobre la visualització de les identitats de gènere a la Viquipèdia.

Taula 12. Objectius específics de la proposta

Objectius específics
Els objectius específics estan relacionats amb si s'analitza o avalua el sistema d'organització del coneixement de Viquipèdia o de Wikidata
Objectius específics per a la diagnosi de la Viquipèdia: a. Anàlisi del sistema d'organització de coneixement per categories [5.1.] b. Avaluació del sistema de categories [5.2.] c. Anàlisi dels processos de presa de decisions [5.3.]
Objectius específics per a la diagnosi de Wikidata: a. <u>Anàlisi del sistema d'organització del coneixement per ontologies</u> [6.1.] b. Avaluació del sistema d'ontologies per a la navegació i cerca [6.2.] c. Anàlisi dels processos de presa de decisions [6.3.]
Conclusions
a. Presència/absència de biaix de gènere a Viquipèdia [5.4.] b. Presència/absència de biaix de gènere a Wikidata [6.4.] c. Proposta de sistema d'organització del coneixement [7.]

5. Diagnosi de Viquipèdia

En aquesta secció de l'informe s'analitza el Sistema d'Organització del Coneixement (SOC) que s'aplica a Viquipèdia per a dues funcions fonamentals: la categorització dels seus continguts, i la cerca mitjançant diferents modalitats i serveis. La nostra atenció se centrarà, específicament, en la versió catalana de Wikipedia, la Viquipèdia.

En primer lloc, se situa el SOC de Wikidata en el marc de la diversitat de sistemes existents, i s'identifiquen els seus components fonamentals a partir de la confrontació amb els estàndards i les bones pràctiques.

La secció es completa amb l'avaluació del SOC mitjançant dos mètodes complementaris: test amb persones usuàries (lectores i editores), i test heurístic elaborat a partir de l'observació de persones expertes (avaluació heurística).

Finalment, s'aporten les conclusions de la presència o absència del biaix de gènere a partir d'aquesta diagnosi.

A l'apartat 6, de manera homòloga, s'analitza el SOC de Wikidata.

5.1. Anàlisi del sistema d'organització del coneixement de Viquipèdia

5.1.1. Categorització del sistema d'organització del coneixement de Viquipèdia

Per al propòsit de categorització i caracterització del SOC aplica en la Viquipèdia, ens situem en la disciplina de l'organització del coneixement. I, dins d'ella, prioritzem el marc d'anàlisi proposat per Zeng (2008). Aquesta autora identifica dues perspectives complementàries: el grau de complexitat de l'estructura del SOC, i les funcions que desenvolupa el SOC.

a. El grau de complexitat de l'estructura oscil·la entre dos extrems: estructura plana (o unidimensional), i estructura multidimensional. En la primera estructura, la plana, els conceptes no presenten cap relació entre ells. En la multidimensional, els conceptes es relacionen entre ells, ja sigui conformant jerarquies basades en el menor o major abast semàntic que representen, o en associacions semàntiques fonamentades en relacions del tipus agent/instrument, causa/efecte, etc.

b. La diversitat funcional també s'organitza en un *crescendo*, amb quatre fites: (i) eliminar l'ambigüitat; (ii) controlar sinònims o equivalents; (iii) establir relacions semàntiques entre termes/conceptes (en particular, relacions jeràrquiques i associatives); i (iv) presentar relacions i propietats dels conceptes en els models de coneixement.

En la majoria dels casos, ambdues perspectives són directament proporcionals. És a dir, com més funcions són assumides per un sistema d'organització del coneixement, major és la seva complexitat estructural.

Pel que fa al grau de complexitat estructural, el SOC de Viquipèdia es compon de categories, és a dir, representació dels conceptes sobre els quals tracten els articles i altres tipus de continguts de l'enciclopèdia. Aquestes categories es relacionen entre si mitjançant una relació semàntica de jerarquia, que vincula categories amb un significat més ampli o general, amb categories amb un significat més reduït o específic. Totes les categories del SOC de Viquipèdia se situen en la jerarquia mitjançant la connexió amb una categoria més general, com a mínim, i amb dues o més categories més específiques. Se n'exceptuen les categories que ocupen el primer nivell de la jerarquia -les anomenades categories temàtiques principals-, i les categories de nivell inferior a

les que es vinculen directament pàgines de la Viquipèdia. La relació semàntica de jerarquia és l'única que estructura de forma completa el SOC de la Viquipèdia.

Pel que fa a les funcions que desenvolupa el SOC de la Viquipèdia, les directrius de categorització exposades en *Viquipèdia:Categorització*³⁴ són clares. La seva funció prioritària és agrupar el coneixement per part de persones editores de l'enciclopèdia, i fer-lo més accessible a les persones lectores, tot facilitant la navegació entre les pàgines. Complementàriament, el SOC permet generar llistes automàtiques sobre els temes relacionats en una matèria.

La confrontació entre l'estructura i les funcions del SOC de Viquipèdia, i la tipologia de sistemes proposats per Zeng (2008), ens permet identificar dos tipus on situar-lo; les taxonomies i els esquemes de categorització. Les primeres són definides com “loosely formed grouping schemes”; les segones com “divisions of items into ordered groups or categories based on particular characteristics.” La clau de la distinció rau en la condició que han de tenir la relació de jerarquia que s'estableix entre les categories; lliure o lleugera en el cas dels esquemes de categorització versus particular o predefinida en el cas de les taxonomies. (En la secció 5.1.2. *Valoració dels principis fonamentals aplicats a la creació de l'esquema de categories de Viquipèdia* fonamentarem i exemplificarem l'adscripció del SOC al tipus d'esquema de categorització).

En el context d'un artefacte digital com és Wikipedia, on la construcció col·laborativa és una de les senyes d'identitat, les dues perspectives anterior indicades no saturen el marc de categorització del SOC. Cal incorporar una tercera perspectiva que tingui en compte quin és el model de presa de decisions pel que fa a la construcció del SOC, d'acord amb la seva estructura i les seves funcions. I situats en aquesta tercera perspectiva de la caracterització del SOC de Viquipèdia, un principi que cal tenir en compte és que qualsevol viquipedista pot impulsar la creació d'una nova categoria. De forma ràpida, podem vincular aquesta possibilitat amb les folksonomies, on la base (genus) de la seva definició incorpora precisament la “categorització col·laborativa”³⁵. La folksonomia a banda de l'aspecte col·laboratiu es complementa amb quatre característiques més (Yedid, 2013):

- Les *folksonomies* es conformen sempre en entorns digitals i d'accés Web.
- Estan basades en l'aplicació d'etiquetes en llenguatge natural, no controlat. I en aquesta falta de control, hem d'afegir-hi la falta d'una estructura semàntica no unidimensional (per exemple amb una mínima relació de jerarquia).
- Les etiquetes són assignades per les persones usuàries dels recursos.
- Finalment, sembla que les *folksonomies* sorgeixen especialment en un entorn digital particular, que és l'entorn social. Les *folksonomies* impliquen el “compartir”, només poden existir si hi ha la cooperació entre els usuaris. Per a molts autors, el poder de les *folksonomies* està connectat a l'acte de l'agregació, no simplement a la creació d'etiquetes. Sense l'entorn social distribuït que permet l'agregació, les etiquetes són simplement paraules planes, només significatives per a la persona usuària que les va triar (Quintarelli, 2005).

³⁴ Wikimedia. (2022, maig 29). *Viquipèdia:Categorització*.

<https://ca.wikipedia.org/w/index.php?title=Viquip%C3%A8dia:Categoritzaci%C3%B3&oldid=30240686>

³⁵ Així s'inicia la definició de *folksonomia* en la Viquipèdia: Folksonomia. (2023). En *Viquipèdia, l'enciclopèdia lliure*. <https://ca.wikipedia.org/w/index.php?title=Folksonomia&oldid=31773340>

Taula 13. Resum de la caracterització en 2 entorns de les *folksonomies*

	Entorn Viquipèdia	Entorn Social Media (<i>P. ex. Flickr o Twitter</i>)
1 Context digital	Sí	Sí
2 Llenguatge natural no controlat	Sí	Sí
3 Llenguatge natural sense jerarquia	No	Sí
4 Assignació d'etiquetes per part de les persones usuàries	Sí	Sí
5 Sense diferències de rol entre els col·laboradors	No. Existeixen diferències de rols entre persones editores i administradores.	Sí
6 Agregació de conceptes sense la possibilitat de corregir o eliminar aportacions anteriors	Es compleix només en el cas de les persones lectores o editores. No es compleix en el cas de les persones administradores. Poden iniciar processos de deliberació per prendre decisions sobre el manteniment o esborrat de categories.	Sí

D'aquestes quatre característiques, l'esquema de categories de Viquipèdia comparteix la primera de forma clara, està en un entorn digital. Val a dir, però, que és la menys idiosincràtica de les quatre, ja que, en el moment actual, disposem de molts SOC nadius digitals.

En canvi, no compleix, en absolut, ni la segona ni la tercera (Vegeu Taula 13). No compleix la segona (llenguatge natural no controlat) perquè l'esquema de categories de Viquipèdia sí que aplica tècniques de control del vocabulari, tant en la designació de les categories, com en la seva estructuració. I no compleix la tercera (llenguatge natural sense jerarquia) perquè qui assigna les categories i, naturalment, qui estableix pautes, criteris, etc. en el seu desenvolupament són persones editores ("creadors de continguts") amb diferents graus d'autoritat. I aquesta tercera característica enllaça, necessàriament, a la matisació de la quarta i darrera característica. És evident que l'entorn de creació i assignació de categories de Viquipèdia es pot qualificar de "social". Però, més enllà del nexa superficial, existeixen profundes diferències si comparem el procés de presa de decisions propi de les *folksonomies* i l'etiquetatge, i el procés de presa de decisions de l'esquema de categories de Viquipèdia i de la categorització en aquest àmbit.

5.1.2. Descripció de l'esquema de categorització

En la secció anterior (5.1.1.) hem situat el SOC de Viquipèdia en la posició que ocupen els esquemes de categorització i les taxonomies pel que fa a la seva estructura i les funcions que desenvolupa. Hem desestimat la qualificació com a folksonomia, tot i la modalitat distribuïda com es construeix. I en la disjuntiva entre aquells dos tipus, hem avançat una major aproximació als primers (esquemes de categories) que als segons (taxonomies). És ara el moment de posar en relleu les distàncies que separen el SOC de Viquipèdia de les taxonomies, per tal de ponderar, a través de la comparació, els trets idiosincràtics com a esquema de categorització.

Amb aquesta finalitat, sotmetrem el SOC de la Viquipèdia a una sèrie de preguntes clau sobre els seus components i la seva estructura. Aquestes preguntes estan basades en les que GGPT (2009) considera que s'han de plantejar en desenvolupar el procés de construcció d'una taxonomia. Les nostres preguntes les agrupem en tres dimensions:

- La identificació dels conceptes a incloure en el SOC.
- Els criteris de classificació aplicats per a la generació de jerarquies.
- Els tipus de relacions semàntiques entre els conceptes.

És en la resposta a aquestes qüestions que caracteritzarem l'esquema de categorització de Viquipèdia, posant en relleu dues consideracions: els trets que l'aproximen i el distancien de les taxonomies, i els beneficis i riscos que aquestes diferents posicions comporten; i els trets que l'alineen i el separen de la perspectiva de gènere.

5.1.2.1. Identificació dels conceptes a incloure en el sistema d'organització del coneixement

Les preguntes que ens fem en aquest apartat són: Com es justifica la incorporació de categories en el SOC? Quin nombre de conceptes es combinen en la formació de categories, és a dir, quina és la seva complexitat? Com es designen les categories?

Les directrius i ajudes que donen suport a les persones editores prioritzen dos criteris en la **justificació de la incorporació de noves categories**; la freqüència en els articles, i la freqüència d'ús. A priori, ambdós criteris són objectivables. Ho podem observar en les següents referències.

- Dins les directrius de Viquipèdia: Categorització³⁶, en l'apartat “Quan cal apuntar a una categoria nova”, s'estableix la següent regla: “En el cas que veiem que una categoria té molts articles (**15 o més**) i que alguns d'ells (**obligatòriament 5 o més**) pertanyen a una àrea més particular d'aquella branca del saber, es pot obrir una nova categoria per contenir aquests articles, però hem de recordar d'incloure la categoria creada dins de la categoria original d'on hem tret els articles nominats i moure els articles relacionats a la nova categoria [...]”.
- Dins la proposta de política Wikipedia: Propuesta de política de categorización³⁷, en les consideracions generals sobre la categorització, s'estableix el principi “No cree categorías de poco uso”.

³⁶ Wikimèdia. (2022). Viquipèdia: Categorització. En *Viquipèdia, l'enciclopèdia lliure*.

<https://ca.wikipedia.org/w/index.php?title=Viquip%C3%A8dia:Categorizaci%C3%B3&oldid=30240686> És una directriu.

³⁷ Wikimedia. (2022). Wikipedia: Propuesta de política de categorización. En *Wikipedia*.

https://es.wikipedia.org/w/index.php?title=Wikipedia:Propuesta_de_pol%C3%ADtica_de_categorizaci%C3%B3n&oldid=142571261. És una proposta de política, convenció, o procés de Wikipedia, i encara no està traduïda al català. També està present en Wikipedia: Categorización de personas. (2023). En *Wikipedia*.

https://es.wikipedia.org/w/index.php?title=Wikipedia:Categorizaci%C3%B3n_de_personas&oldid=150335316#Criterios_de_categorizaci%C3%B3n

Fixem-nos que la justificació orientada a la freqüència d'aparició s'estableix en dues xifres concretes. Categories assignades a 15 o més articles, i dels quals 5 o més tracten sobre un concepte més específic, constitueix el barem precís per a generació d'una nova categoria.

Pel que fa a la **complexitat de les categories**, en l'esquema de Viquipèdia trobem representació dels dos extrems del contínuum.

- Categories que representen un sol concepte. Per exemple, *Infermeria* (representa una àrea de coneixement) o *Infermers* (representa una activitat professional).
- Categories que representen combinacions de dos o més conceptes. Per exemple, *Persones de l'àmbit catalanoparlant per origen i activitat* (representa la combinació d'entitat + espai + origen + activitat), *Bibliotecaris catalans del sud contemporanis* (representa la combinació d'activitat professional + origen + temps), *Infermers catalans del sud contemporanis* (representa la combinació d'activitat professional + origen + temps), *Mestres d'educació primària catalans del sud contemporanis* (representa la combinació d'activitat professional + origen + temps).

La primera solució s'alinea amb el model de **postcoordinació**, que dona a la persona usuària el poder de selecció i combinació de conceptes en sistemes de cerca per recuperació. La segona solució, en canvi, s'alinea més amb el model de **precoordinació**, i, tradicionalment, s'ha orientat a facilitar els processos de cerca per exploració i/o navegació. L'aplicació d'aquesta segona solució té diversos riscos.

- Incrementa les dimensions del sistema d'organització de coneixement, de vegades, fins a límits inabastables, si no s'estableixen límits en les possibilitats combinatòries dels conceptes.
- Allunya el conjunt del sistema de la connexió necessària amb la intuïció dels usuaris, si la precoordinació no es fonamenta en la freqüència de cerca dels usuaris.

Val a dir, però, que la identificació tradicional entre precoordinació i sistemes de cerca per exploració i navegació no es pot mantenir en el context de les interfícies digitals actuals, on ja no és necessari (ni òptim) l'itinerari seqüencial per les cadenes de les jerarquies de conceptes, fins a arribar al terme que millor respon una necessitat d'informació.

En conjunt, l'abús de la precordinació desdibuixa el protagonisme de la persona usuària en les decisions de combinació de conceptes per atendre les seves necessitats i expectatives en moments puntuals, i en períodes llargs. A la Viquipèdia, l'absència de directrius sobre aquest aspecte fa pensar en decisions espontànies pel que fa a la disjuntiva precoordinació/postcoordinació, i, a més, en una conseqüència perniciosa del fenomen de la "falsa facetació" a què farem referència més endavant.

Una tercera qüestió lligada a la identificació de les categories a incloure en el SOC és la designació de les categories. Es justifica a partir de resultats del sistema de presa de decisions propi de Viquipèdia. En el cas de les biografies, la decisió presa ha privilegiat fins a 16 criteris de categorització (per contrast, tan trivial com ara "Innovadors del sector de la motocicleta", o tan ambigu com ara l'"estatus" que des d'una perspectiva més interseccional Rodó-Zárata (2021) ens posa en qüestió el perquè del rebuig del gènere en la divisió de categories. Això voldria dir que el gènere es pot emprar com a categoria general però no per subcategoritzar. Actualment, a la Viquipèdia la categoria general sempre empra el gènere masculí, excepte en casos molt específics com *Llevadores* o *Dones Barbudes*. Per tant, el gènere per defecte a les categories de primer nivell és el masculí i també en les subcategories, ja que no es deixa dividir per gènere en aquest nivell tal com indica la directriu i es va corroborar en la votació, excepte casos especials com els comentats.

La directriu Viquipèdia: Categorització³⁸, de l'any 2018 (abans de la votació de les categories de gènere del 2022)³⁹ justifica aquesta regla sobre la base de la manca de consens, i fa explícita la diferència de criteri respecte a altres wikipèdies. En l'apartat “4.3. Anàlisi de la creació i processos de presa de decisions dels sistemes d'organització del coneixement de Viquipèdia: categories”, aprofundim sobre el context en què la comunitat de Viquipèdia ha pres decisions com aquesta, i s'assenyala com, des del punt de vista de la justificació en la selecció i designació de conceptes, ens trobem davant d'un criteri arrelat en biaix cultural. Els resultats són que l'etiquetatge de les categories relatives a persones, s'aplica exclusivament el gènere masculí; tant si la categoria es focalitza directament en un col·lectiu de persones (Científics, Il·lustradors científics, Crítics culturals, Promotors culturals, Directors de tecnologia...) com si el col·lectiu forma part d'una designació més complexa (Congrés Internacional de Matemàtics). I resulta difícil de justificar la incorporació de la categoria *Infermers*⁴⁰, que inclou sis pàgines, totes elles relatives a persones del gènere femení.

5.1.2.2. Criteris de classificació aplicats en la generació de jerarquies

Les preguntes que ens fem en aquest apartat són dues: Quins criteris s'utilitzen per a la classificació dels objectes del SOC? Quantes categories d'abast semàntic més ampli (supercategories) es poden vincular a una categoria? S'aplica l'anàlisi de facetes com a principi d'estructuració general de les categories del SOC?

En general, la classificació d'objectes en un SOC pot respondre a diferents tipus de relació jeràrquica, cada una de les quals correspon a situacions lògiques diferents. La norma UNE-ISO 25964-1⁴¹ en diferencia tres:

- La **relació genèrica**, consistent en la connexió entre una classe o categoria i els seus membres o espècies.
- La **relació tot-part**, que cobreix un rang limitat de situacions en les quals una part d'una entitat o sistema pertany únicament a un tot particular que la posseeix. La norma UNE-ISO 25964-1 n'identifica quatre: sistemes o òrgans del cos; localitzacions geogràfiques; disciplines o camps del coneixement; estructures socials jerarquizades. Altres estàndards, com GGPT (vegeu la nota al peu de la Taula 9) les amplia substancialment.
- La **relació enumerativa** connecta un concepte general, com una classe de coses o esdeveniments, i una instància individual de la classe esmentada, la qual sovint es representa amb un nom propi.

Amb independència del tipus de relació seleccionada, els estàndards i bones pràctiques recomanen garantir la prova de consistència consistent en la generació de cadenes de jerarquia. Aquesta prova consisteix en el fet que tots els conceptes (categories) del nivell més alt, fins al nivell més baix, pertanyen a la mateixa categoria fonamental, com són objectes, materials, persones, accions, llocs, temps, etc. És una garantia de la disjunció semàntica dels conceptes representats; és a dir, no hi haurà dos conceptes amb encavalcament semàntic.

D'altra banda, les taxonomies han singularitzat la relació genèrica com l'única aplicable, on tot objecte capturat pel concepte més general és també capturat pel concepte més específic. Aquesta relació es pot expressar com una relació “és un”; un membre d'*Infermers de Catalunya*

³⁸ Wikimèdia. (2022). Viquipèdia: Categorització. En *Viquipèdia, l'enciclopèdia lliure*.

<https://ca.wikipedia.org/w/index.php?title=Viquip%C3%A8dia:Categoritzaci%C3%B3&oldid=30240686>. És una directriu.

³⁹ El resultat de la votació no va modificar el que ja deia la directriu de no emprar el criteri de gènere per a la divisió de categories

https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Presa_de_decisions/2018/Sobre_la_categoritzaci%C3%B3_per_g%C3%A8nere

⁴⁰ Categoria:Infermers. (2021). En *Viquipèdia, l'enciclopèdia lliure*.

<https://ca.wikipedia.org/w/index.php?title=Categoria:Infermers&oldid=26858616>

⁴¹ Publicada com AENOR. (2014). *UNE-ISO 25964-1: Información y documentación: Tesoros e interoperabilidad con otros vocabularios. Parte 1: Tesoros para la recuperación de la información*. AENOR.

és un membre d'*Infermers*. L'aplicació de la relació genèrica enfront de, per exemple, la relació part-tot, exigeix el compliment d'una regla de validesa lògica addicional; la prova "tots-i-alguns". Tots els infermers catalans són infermers; alguns infermers són infermers catalans. D'aquesta forma, es garanteix l'aplicació de la propietat transitiva en tots els nivells d'una cadena jeràrquica. El concepte específic és un (membre o tipus) del seu concepte general; aquest és un (membre o tipus) del seu concepte general...; i així fins al concepte més ampli de la cadena de jerarquia.

L'aplicació d'aquests principis són coherents amb la utilitat que s'espera de l'aplicació de les jerarquies a la cerca d'informació. D'una banda, la persona usuària pot navegar des dels nivells més baixos fins als més alts, o viceversa, amb la seguretat que mai perdrà el control sobre els conceptes que s'espera trobar en els nivells més alts o més baixos. I tampoc el perdrà quan la modalitat de cerca utilitza el SOC per a l'expansió de la cerca; és a dir, la incorporació automàtica de conceptes específics en l'equació de consulta.

No compleixen les proves lògiques indicades altres tipus de relacions com és la relació tot-part. I molt menys les que atenen a associacions entre conceptes com són causa-efecte, agent-instrument, àrea del coneixement-objecte d'estudi, etc. En tots els casos, les jerarquies generades atenen a criteris pragmàtics, però s'allunyen de la intuïció i, sovint, de la comprensió dels usuaris.

El criteri de classificació aplicat en l'esquema de categorització de Viquipèdia no respon, de forma unívoca, a la relació genèrica. En la versió en castellà de les directius Wikipedia: Categorización⁴² la manca d'aquest requeriment es sintetitza en l'expressió "Falta de transitividad conceptual", i es defineix de la següent forma:

"La regla principal es que cada subcategoría debe ser más específica que las categorías en las que está incluida. Sin embargo, por la naturaleza tan amplia y diversa de las categorías, la contención conceptual puede no heredarse a subcategorías de nivel más profundo. Esto no constituye un error."

La barreja de criteris de classificació es pot observar, clarament, en la subdivisió aplicada a les professions feminitzades que observem de forma prioritària en aquest informe.

Infermeria

Infermers
Llevadores
Primers auxiliis

Biblioteconomia

Associacions professionals relacionades amb la informació i la documentació
Bibliografia
Bibliometria
Biblioteca de Catalunya
Bibliotecaris
Biblioteques
Centres de documentació
Directoris
Escola de Bibliotecàries
Facultat de Biblioteconomia i Documentació (UB)
Presidents de l'IFLA

⁴² Wikipedia: Categorización. (2022). En *Wikipedia*.

https://es.wikipedia.org/w/index.php?title=Wikipedia:Categorizaci%C3%B3n&oldid=147798114#Falta_de_transitividad_conceptual

Educació primària

Alumnes per centre d'educació primària

Centres d'educació infantil i primària

Mestres d'educació primària

El risc de no respectar la relació genèrica en l'establiment de la relació de jerarquia entre categories és clar: les cadenes jeràrquiques generades dificulten als usuaris comprendre l'estructura global de l'estructura del SOC, i resulten poc intuïtives i difícils d'aprendre i recordar en processos d'exploració dels conceptes i d'expansió de les cerques.

La segona pregunta que ens hem plantejat en aquest apartat (*Quantes categories d'abast semàntic més ampli (supercategories) es poden vincular a una categoria?*) ens condueix a la distinció entre **monojerarquia** i **polijerarquia**. Un SOC és monojeràrquic quan cada concepte (excepte el concepte superior) té un i només un concepte més ampli. En cas contrari, quan els conceptes poden tenir més d'un concepte superior, ens trobem davant un SOC polijeràrquic. En aquest segon cas, els termes més amplis no són disjunts, sinó que es superposen. Una situació possible en la polijerarquia és quan un concepte és específic d'un de superior a partir de la relació genèrica, i d'altres generats a partir d'altres tipus de relacions, com la part-tot.

Les GGPT (vegeu nota al peu de la Taula 9) qualifica la monojerarquia com una tècnica que aporta menor complexitat i, també, menor expressivitat als SOC; mentre que la polijerarquia actua en sentit contrari pel que fa a aquests dos valors. La polijerarquia proporciona diverses vies per arribar a un mateix contingut, i aquesta diversitat es trasllada també al sistema de cerca. No obstant això, els SOC polijeràrquics tendeixen a ser difícils d'entendre, en particular, quan totes les classificacions o distincions possibles de termes s'implementen en paral·lel.

En l'esquema de categorització de Viquipèdia, l'aplicació de polijerarquia és força habitual. A més, en els casos que s'aplica, el nombre de supercategories és superior a dues. En són exemples les categories corresponents a les professions feminitzades (personal bibliotecari, personal d'infermeria, i personal d'ensenyament primari i secundari) que presentem a continuació⁴³. (Vegeu la Imatge 6. Arbre de categories de Viquipèdia).

- La categoria *Bibliotecaris* té com a categories superiors *Biblioteconomia*, *Categories amb enllaç commonscat des de Wikidata*, *Gestors d'informació* i *Professions de la literatura*.
- La categoria *Infermers* té com a categories superiors *Biografies per activitat*, *Categories amb enllaç commonscat des de Wikidata* i *Infermeria*.
- La categoria *Mestres d'educació primària* té com a categories superiors *Categories amb enllaç commonscat des de Wikidata*, *Educació primària* i *Professors per nivell educatiu*.

La polijerarquia s'aplica també a categories creades a partir de precoordinaió de més d'un concepte, com es pot observar en els següents exemples.

- La categoria *Persones de l'àmbit catalanoparlant per origen i activitat* té com a categories superiors *Europeus per origen i activitat*, *Persones de l'àmbit catalanoparlant per activitat*, i *Persones de l'àmbit catalanoparlant per origen*.
- La categoria *Bibliotecaris catalans del sud contemporanis* té com a categories superiors *Bibliotecaris catalans*, *Bibliotecaris espanyols* i *Catalans del sud contemporanis per activitat*.

⁴³ El terme *commonscat* fa referència a les categories amb què es classifiquen les imatges de Wikimedia Commons.

- La categoria Infermers catalans del sud contemporanis té com a categories superiors *Catalans del sud contemporanis per activitat, Infermers catalans i Infermers espanyols*.

La combinació de precoordiació i polijerarquia en les categories que acabem de citar té un efecte multiplicatiu pel que fa a la seva complexitat semàntica, i, en els contextos de cerca, a les exigències dels usuaris pel que fa a esforços de comprensió, aprenentatge i memorització. I aquestes exigències tampoc es redueixen dràsticament encara que la polijerarquia s'apliqui a categories que representen un únic concepte.

En els casos on un mateix concepte pot ser observat a partir de diferents perspectives -com possibilita la polijerarquia- i existeix la necessitat de combinar diferents conceptes per a representar de forma exhaustiva molts dels continguts classificats -com possibilita la precoordiació-, una alternativa d'estructuració global del SOC és la classificació facetada. Les facetes són agrupacions de conceptes que representen un, i únicament una característica de divisió d'una àrea del coneixement, o del coneixement en el seu conjunt. Les facetes són mútuament excloents entre sí, i el conjunt de les facetes permeten representar de forma completa el coneixement. Existeixen propostes de facetes d'aplicació general, per a tots els dominis del coneixement; les 10 categories fonamentals d'Aristòtil, la fórmula PMEST de Ranganathan, o, més recentment, els 9 tipus de conceptes proposats per la norma *UNE-ISO 25964-1*⁴⁴.

- Objectes i les seves parts físiques
- Materials
- Activitats o processos
- Esdeveniments o ocurrències
- Propietats de persones, coses, materials o accions
- Disciplines o camps temàtics
- Unitats de mesura
- Tipus de persones i organitzacions
- Entitats individuals (representades per noms propis)

La Viquipèdia compleix tres dels escenaris recomanables per a l'aplicació del model de classificació facetada segons les GGPT (vegeu nota al peu de la Taula 9).

- Cobreix àrees interdisciplinàries amb més d'una perspectiva per mirar objectes de contingut o necessitat de combinació de conceptes.
- Té un SOC amb múltiples jerarquies però límits poc clars.
- El SOC s'orienta a la classificació d'objectes digitals on la ubicació i la col·locació no tinguin importància.

I se'n podria beneficiar de dues avantatges claus que destaca aquestes directrius.

- Una classificació facetada té unes dimensions molt més reduïdes que una de mono- o polijerarquica, sense perdre la seva expressivitat. En conseqüència, és més fàcil de construir i mantenir.
- La classificació per facetes permet als usuaris cercar o navegar per recursos amb més flexibilitat, ja que poden cercar un recurs des de diferents angles.

El SOC de Viquipèdia no fa una aplicació exhaustiva, sistemàtica, del model de classificació facetada. Aquest fet l'allunya d'obtenir els beneficis que identifiquen les GGPT (veure nota al peu de la Taula 9). Dues són les principals desviacions del model de facetació estricta.

⁴⁴ AENOR. (2014). *UNE-ISO 25964-1: Información y documentación: Tesoros e interoperabilidad con otros vocabularios. Parte 1: Tesoros para la recuperación de la información*. AENOR.

- En el nivell bàsic de la classificació no s’ha realitzat una anàlisi de facetes pròpiament dit. Les vuit categories temàtiques principals⁴⁵ que es van establir, de forma definitiva, el 2016, no són excloents entre si.
 - Biografies
 - Ciència
 - Cultura
 - Esdeveniments
 - Humanitats
 - Informació
 - Llocs
 - Tecnologia

Hi ha cinc categories principals que responen a la divisió en disciplines del coneixement humà: Ciència, Humanitats, Informació i Tecnologia. El criteri aplica per a la segmentació de les disciplines, però, dona pas a múltiples solapaments en els nivells subsegüents. És el cas, per exemple, de la categoria *Estudis culturals*.

En canvi, Esdeveniments i Llocs semblen generades a partir de categories fonamentals, mentre que Biografies apunta directament a un gènere documental. Totes elles són aplicables a totes les disciplines del coneixement humà. És a dir: la Ciència la podem subdividir per àrees o disciplines científiques en diversos nivells jeràrquics d’especificació, però també la podem subdividir pels llocs de la Ciència, els Esdeveniments de la Ciència i les Biografies de la Ciència.

Finalment, la categoria Cultura no es pot identificar, clarament, com a disciplina o com a fenomen.

- En la subdivisió de les categories temàtiques principals, podem distingir tres orientacions diferents.
 - Subcategorització orientada a la subfacetació.
 - Biografies (8 cat., 29 p.). Es concentren en la subcategoria “Categories de biografies per paràmetre”.
 - Esdeveniments (9 cat., 5 p.). Les facetes es concreten en sis de les nou subcategories: “Esdeveniments en curs”, “Esdeveniments per mes”, “Esdeveniments per segle”, “Esdeveniments per tema”, “Esdeveniments per territori” i “Esdeveniments per tipus”.
 - Subcategorització orientada a l’especificació de disciplines.
 - Humanitats (13 cat., 12 p.). Inclou: “Administració”, “Art”, “Dret”, “Estudis culturals”, “Filologia”, “Filosofia”, “Humanitats mèdiques”, “Lingüística i Teologia”.
 - Informació (5 cat., 1 p.). Inclou: “Ciències de la informació”, “Comunicació” i “Periodisme”.
 - Les categories temàtiques principals d’aquest tercer grup es caracteritzen per incorporar un gran de nombre de subcategories.

⁴⁵ Categoria:Classificacions temàtiques principals

https://ca.wikipedia.org/wiki/Categoria:Classificacions_tem%C3%A0tiques_principals

Ciència (41 cat., 70 p.)

Cultura (32 cat., 54 p.)

Tecnologia (22 cat., 273 p.)

Entre aquestes subcategories trobem, d'una banda, les que podrien conformar subfacetes, incloent-hi la subfaceta de disciplina. Per exemple, en el cas de Cultura, trobem de forma explícita les subfacetes Cultura per grup humà i Cultura per territori. Les altres 30 es podrien simplificar de la següent forma.

Faceta de persones, on se situarien Crítics culturals, Promotors culturals...

Faceta de processos i fenòmens, on se situarien Cultura de l'anarquisme, Cultura LGBT...

Faceta d'esdeveniments, on se situarien Esdeveniments culturals...

Faceta d'objectes, on se situarien Indumentària i Obres.

Faceta de disciplines, on se situarien Art, Història cultural i Religió.

- La categoria principal Llocs manifesta un tractament singular, a mig camí entre les dues orientacions anteriors.

5.1.2.3. Relacions semàntiques entre els conceptes

La pregunta que ens formulem en aquest apartat és Quins tipus de relacions entre els objectes s'han de triar per incloure'ls a la taxonomia?

L'única relació semàntica que s'aplica en l'esquema de categories de Wikipedia és la de jerarquia, és a dir, la relació entre un parell de conceptes dels quals un té un abast que queda completament dins l'abast de l'altre. Resten fora de l'estructura, doncs, altres dues relacions que, quan s'apliquen, juguen un rol fonamental en el suport a la persona usuària en la cerca per exploració: la **relació d'equivalència** i la **relació d'associació**.

La **relació d'equivalència** s'estableix entre dos termes quan ambdós representen el mateix concepte. L'origen de l'equivalència pot estar motivada per la sinonímia pròpia del llenguatge natural, o pot estar motivada per identificacions entre conceptes que, si bé en el llenguatge natural no són pròpiament sinònims, es tracten com a tals en el context del SOC per a reduir el desbordament del vocabulari tant en la dimensió de profunditat, com en la dimensió de l'amplada.

Els estàndards i bones pràctiques recomanen la incorporació d'aquesta relació semàntica en SOC complexos com és l'esquema de categories de Viquipèdia. Cal, això sí, designar un dels termes equivalents com a preferent, per tal de prioritzar-lo en contextos on és difícil la concurrència de totes les formes. Des de la perspectiva de l'usuari, la relació d'equivalència s'alinea amb la visibilització de la diversitat. Diferents designacions d'un mateix concepte, sovint vinculades a perspectives diverses, poden concórrer en els processos d'exploració i recuperació de continguts. A més, és possible vehicular mecanismes de personalització del vocabulari per a cada persona usuària en funció de les seves preferències.

Algunes indicacions sobre els aspectes formals de les designacions de les categories, si bé no són exhaustives, es pot consultar en *Viquipèdia: Categorització*⁴⁶. I les convencions sobre la designació de títols amplien el seu abast a les categories, amb importants especificacions. Efectivament, de forma més explícita, la versió en castellà de *Wikipedia: Convenciones de títulos*⁴⁷ estableix:

“Las categorías se crean de manera análoga a los artículos, siempre respetando las convenciones de títulos; sin embargo, todo su contenido es opcional, salvo por la identificación de su o sus categorías madre.”

Una de les especificitats en la designació de les categories és, precisament, l'exclusió de la recomanació de fer redireccions⁴⁸; exclusió que tanca la porta a la relació d'equivalència en l'esquema de categories de Wikipedia.

La **relació d'associació** s'estableix entre un parell de conceptes que no estan relacionats jeràrquicament, però que comparteixen una forta connexió semàntica. Els estàndards i bones pràctiques destaquen que la incorporació d'aquesta relació enhance the usefulness and expressivity of a taxonomy by suggesting additional (associated) terms for use in indexing and retrieval and of named relationships in particular. També comporta riscos, ja que s'incrementa la complexitat del SOC en la seva construcció i manteniment. En tot cas, l'establiment d'associacions entre conceptes no pot ser espontània ni pot respondre a visions personals. És necessari l'establiment de patrons de relacions entre conceptes que s'han d'aplicar en el conjunt del SOC, o en seccions de l'esquema.

5.2. Avaluació del sistema de categories per a la navegació i cerca

L'avaluació que plantegem del sistema de navegació i cerca a través de categories i dels diferents cercadors de Viquipèdia es planteja des del focus en la usabilitat del sistema, és a dir la facilitat d'aprendre'l i usar-lo, així com de l'experiència de la persona usuària a partir de la interacció, és a dir si és satisfactori, motivador, etc.

D'una banda, l'avaluació s'ha planejat amb experiments amb persones usuàries proporcionant dades quantitatives, però sobretot de caràcter qualitatiu. Amb el test amb usuaris es comprova si les interfícies i els processos estan ben dissenyats i si els usuaris troben obstacles i dificultats per tal que les persones usuàries puguin assolir els seus objectius a la plataforma digital (Preece et al., 2019). D'altra banda, l'avaluació heurística consisteix en l'estudi i l'avaluació de la interfície per un expert en usabilitat o més d'acord amb un conjunt de regles i principis d'usabilitat establerts prèviament. Aquests principis d'usabilitat que serveixen de base per a l'avaluació es denominen heurístics (Preece et al., 2019).

5.2.1. Avaluació amb persones usuàries

Les proves amb persones usuàries han tingut com a objectiu avaluar els sistemes de navegació i cerca que la Viquipèdia posa a disposició així com la perspectiva que posseïen els usuaris respecte a l'organització categòrica de l'enciclopèdia col·laborativa catalana (Vegeu taula 7 per

⁴⁶ Wikimèdia. (2022). *Viquipèdia: Categorització*. En *Viquipèdia, l'enciclopèdia lliure*.

<https://ca.wikipedia.org/w/index.php?title=Viquip%C3%A8dia:Categoritzaci%C3%B3&oldid=3024068>. És una directriu.

⁴⁷ Wikimèdia. (2023). *Wikipedia: Convenciones de títulos*. En *Wikipedia*.

https://es.wikipedia.org/w/index.php?title=Wikipedia:Convenciones_de_t%C3%ADtulos&oldid=151323697. És una ajuda.

⁴⁸ Wikimèdia. (2023). *Viquipèdia: Anomenar pàgines*. En *Viquipèdia, l'enciclopèdia lliure*.

https://ca.wikipedia.org/w/index.php?title=Viquip%C3%A8dia:Anomenar_p%C3%A0gines&oldid=31193386. És una directriu.

veure els experiments dels tests d'usuaris). Com hem exposat anteriorment, hem triat les variables d'anàlisi que més s'adeqüen al nostre estudi d'usabilitat.

5.2.1.1. Eficiència

S'han calculat les mitjanes de clics de tots els usuaris i els clics del recorregut cognitiu fet per un expert en què estableix el llindar mínim de clics, i de forma global es presenta a la taula 14.

Taula 14. Eficiència de cerca i navegació a les diferents opcions de Viquipèdia en comparació a la probabilitat de retorn

Nombre de la tasca	Mitjana clics fets (usuaris)	Clics mínims necessaris (recorregut cognitiu)	Eficiència de navegació	Mitjana de retorns	Probabilitat de retorn
Tasca 1: navegació autònoma	5,9	3	51%	0,7	12%
Tasca 2: cercador general	9,1	3	33%	0,9	10%
Tasca 3: cercador de gènere	7,7	4	52%	1,1	14%
Tasca 4: cerca interna	7,2	7	97%	1,1	15%
Tasca 5: intersecció	8,4	3	36%	0,7	8%
Tasca 6: índex i arbre de categories	6,1	2	33%	0,5	8%
Tasca 7: índex de categories	7,9	5	63%	1,1	14%
Tasca 8: arbre de categories	13,4	6	45%	2,1	16%

En primer lloc, s'han proporcionat dades sobre la mitjana de clics realitzats per completar cada tasca específica. Aquesta mètrica ofereix una indicació de la quantitat de passos necessaris per arribar a la informació desitjada. Un menor nombre de clics indica una interacció més eficient, ja que la persona usuària pot accedir més ràpidament a la informació que necessita. Com que la diferència entre els usuaris lectors i les persones editores no ha sigut significativa a l'hora d'establir la mitjana de clics fets, hem fet un recompte global de l'eficiència de la recuperació de la informació a Viquipèdia (Vegeu la columna de "Mitjana clics, usuaris" de la Taula 14).

A continuació s'ha inclòs els resultats d'aplicar la tècnica del recorregut cognitiu, és a dir, el nombre mínim de clics necessaris per a cada tasca, que representa el nombre òptim de passos per completar amb èxit la tasca. Aquesta xifra recordem que serveix com a referència per avaluar si els usuaris estan aconseguint una eficiència semblant o superior al mínim requerit. Quan el nombre de clics realitzats s'apropa al nombre mínim, es pot considerar que l'experiència de la persona usuària és més eficient. (Vegeu la columna de "Clics mínims necessaris, expert en el recorregut cognitiu" de la Taula 14).

La columna "Eficiència de navegació" (Vegeu Taula 14) indica el percentatge d'eficiència amb què els usuaris han completat cada tasca en comparació amb els clics mínims necessaris establerts en el recorregut cognitiu. Aquesta mètrica proporciona una perspectiva relativa de la facilitat o dificultat amb què els usuaris han estat capaços de navegar per la Viquipèdia i obtenir la informació desitjada. Quan l'eficiència de navegació és alta, indica que els usuaris han pogut completar la tasca de manera més eficient, apropant-se als clics mínims requerits.

Finalment, la probabilitat de retorn fa referència a la freqüència en què els usuaris han hagut de tornar a pàgines anteriors durant la realització de la tasca. Un percentatge baix indica que les persones usuàries han pogut progressar de manera lineal i sense contratemps, mentre que un percentatge alt pot indicar dificultats o necessitat de revisar passos anteriors per arribar a satisfer la necessitat informativa.

En global, pel que fa a la cerca, destaca de manera molt positiva l'eficiència del cercador intern amb una taxa del 97%. Amb tot, és important contextualitzar aquestes dades amb la probabilitat de retorn, on la cerca interna presenta poca eficiència (15% de taxa de retorn) la tasca amb una probabilitat més alta de què la persona usuària hagi de tornar a pàgines anteriors durant la seva navegació per aconseguir el resultat desitjat. En canvi, el cercador general és on les persones usuàries han experimentat més dificultats per trobar el resultat desitjat.

I pel que fa a la navegació a través de les categories, presenta més eficiència l'arbre de categories (33%) malgrat la seva taxa de probabilitat de retorn és prou alta (16%).

5.2.1.1. Efectivitat

A partir de la tècnica del *think aloud* (pensament manifest) s'han recuperat dades qualitatives amb les quals s'ha conclòs que les persones usuàries expressen problemes d'usabilitat i dificultats en la recuperació de la informació tant en la navegació per categories com amb els cercadorr. Per tant, les categories i el cercador no són prou efectius des del punt de vista dels usuaris per aconseguir els resultats desitjats. Així, per exemple un usuari ha expressat que "la Viquipèdia és confusa i poc intuïtiva", i fins i tot, hi han coincidit viquipedistes experimentats. Aquesta percepció de complexitat compartida demostra la dificultat que presenta arribar a la informació plantejada en les tasques de l'experiment, i per tant la dificultat de satisfer les necessitats informatives, arribar els resultats desitjats, en l'entorn real.

En relació amb la recuperació de la informació, els usuaris han observat limitacions en aquest aspecte. La Viquipèdia no està específicament dissenyada per a la cerca i recuperació d'informació, fet que pot impactar negativament l'experiència de l'usuari. Alguns usuaris han assenyalat que la recuperació d'informació a través de cercadors externs és més efectiva. Gairebé totes les persones usuàries van confessar que utilitzaven el cercador Google per produir els resultats desitjats a Viquipèdia. Això indica que, en comparació amb aquest cercador o d'altres disponibles en línia, la Viquipèdia pot tenir limitacions en la seva efectivitat per a la cerca d'informació precisa. Alhora aquesta dependència de cercadors externs deixa a l'enciclopèdia en una posició d'indefensió, a mercès dels algoritmes, tal com evidencia l'Aliança per la Presència Digital del Català en el seu informe de posicionament del català als cercadors, de juny de 2023, en què es denuncia que Google ha fet perdre visites a les pàgines web catalanes⁴⁹.

⁴⁹ https://aliancadigital.cat/wp-content/uploads/2023/06/informe_posicionament_v.1.11_CAT.pdf

5.2.1.1. Facilitat d'aprenentatge

Les persones usuàries han valorat la intuïtivitat i l'experiència global de la cerca a Viquipèdia en les tasques fetes. Els valors estan basats en una escala de 0 a 5, on 0 representa "gens intuïtiu" i 5 representa "totalment intuïtiu". Calculant la mitjana, podem obtenir una puntuació mitjana per a la intuïtivitat del cercador a la Viquipèdia.

La mitjana de les puntuacions proporcionades per a la intuïtivitat dels cercadors de la Viquipèdia és de 2.2. Aquesta puntuació indica que, de mitjana, els usuaris consideren que el cercador és moderadament intuïtiu, però encara hi ha una àrea d'optimització per millorar l'experiència de cerca a la plataforma. Alguns usuaris han expressat que troben que el cercador és suficientment intuïtiu per a fer cerques bàsiques, però més complex per a cerques més avançades. Això suggereix que potser hi ha una manca de claredat o eines per a facilitar la cerca de continguts més específics i satisfer les necessitats més sofisticades.

Una persona va destacar les dificultats per trobar les categories dins dels articles de Viquipèdia, la qual cosa complica la navegació per categories perquè no en coneixen l'existència o és gairebé invisible als lectors. Això pot afectar la capacitat dels usuaris per a afinar les seves cerques i accedir a informació més específica. Una altra persona usuària va assenyalar que el cercador no permet l'ús de llenguatge natural, la qual cosa pot limitar la facilitat d'utilització per a aquells que prefereixen introduir termes de cerca de manera més natural o casual.

Les persones participants en els experiments també van esmentar que, tot i que el cercador és intuïtiu, la seva visibilitat pot ser limitada, ja que està ubicat al marge superior esquerra de la pàgina. Això pot dificultar la seva localització per a alguns usuaris, i se suggeria millorar-ne la visibilitat o l'accessibilitat.

En general, els comentaris dels usuaris indiquen que, malgrat la dificultat de saber on trobar el cercador de gènere, un cop s'ha utilitzat, és fàcil d'entendre com funciona i és útil. És important tenir en compte aquests comentaris per millorar l'experiència de cerca i satisfer les necessitats dels usuaris a la Viquipèdia. Això es pot també pensar considerant els resultats de l'eficiència, on a les quatre primeres tasques – que són de navegació amb la Viquipèdia directament, mentre que el cercador d'intersecció utilitza una interfície fora de l'enciclopèdia online –, pot existir una corba d'aprenentatge en què a mesura que la persona usuària va tenint coneixement de la navegació, desenvolupa menys problemes a l'hora de recórrer la Viquipèdia, amb menys necessitat de clics per tasca.

Ara bé, la valoració dels cercadors, la qual es calcula amb la satisfacció qualificada entre 0 "gens" i 5 "totalment" per les persones usuàries i editores, és clarament diferent si és feta per un editor o per un lector, com indica la taula 15:

Taula 15. Valoració mitjana dels cercadors

Valoració	Cercador general	Cercador de gènere	Cerca interna	Cercador d'intersecció
Mitjana	3,15	2,55	2,35	2,15
Mitjana (editors/res)	3,4	2,3	2,6	2,6
Mitjana (lectors/res)	2,5	2	1,9	0,9

Les persones editores, amb un coneixement previ més ampli que els usuaris, han valorat positivament els cercadors, sent el cercador de gènere el que menys. En canvi, per a les persones lectores és el cercador general el que mínimament aprova l'avaluació, com que és el cercador més emprat de tots els avaluats. El cercador de gènere no l'usen perquè no saben com trobar-lo.

5.2.1.1. Trobabilitat

La variable avalua la facilitat de trobar el que la persona usuària necessita. En aquest cas s'ha calculat si la persona usuària era capaç de complir la tasca proposada (Vegeu la taula 16 amb els resultats d'assoliment).

Taula 16. Nombre de persones usuàries que satisfan la necessitat d'informació plantejada en cada tasca

	Tasca1	Tasca2	Tasca3	Tasca4	Tasca5	Tasca6	Tasca7	Tasca8
Mtjana global (%)	60	90	15	45	45	95	100	70
Editors/res (%)	100	100	90	80	80	100	100	80
Lectors/res (%)	20	80	40	10	10	90	100	60

En general, els resultats mostren una variabilitat significativa en les taxes d'èxit per tasca, segons si és realitzada per una persona editora o usuària a la Viquipèdia. A la taula 16 es veu com les persones editores només tenen lleugeres dificultats amb el cercador intern i el de gènere. Mentre que en el cas de les persones lectores la taxa és molt baixa en els dos cercadors, intern i gènere.

El grup de persones editores mostren generalment una taxa d'èxit més elevada en diverses tasques, destacant en tasques com la cerca interna, on tots els editors aconseguen un 100% d'èxit. Això suggereix que el grup d'editors tenen un coneixement més profund de l'estructura i el funcionament de la Viquipèdia, la qual cosa els permet abordar de manera més efectiva les tasques més complexes.

Per contra, el grup de persones lectores mostren taxes d'èxit més variables, amb unes taxes d'èxit força diferent entre tasques. Les tasques que requereixen un nivell més elevat d'expertesa, com la cerca d'intersecció (tasca 5) o la cerca interna o avançada (tasca 4), mostren taxes d'èxit més baixes per a les persones usuàries. Això reflecteix les dificultats que poden experimentar els usuaris sense un coneixement profund de l'estructura i les funcionalitats de la Viquipèdia. En canvi, les persones que són lectores o usuàries de la Viquipèdia es desenvolupen molt bé amb l'índex de categories (tasca 7), amb l'arbre de categories (tasca 6) o amb el cercador general (tasca 2).

En resum, aquests resultats indiquen que els editors tenen una millor comprensió i domini de com s'organitzen els continguts a la Viquipèdia, cosa que els permet obtenir majors taxes d'èxit en les diferents tasques. D'altra banda, els lectors enfronten desafiaments addicionals i mostren nivells variables d'èxit en funció del grau de complexitat de cada tasca.

5.2.1.1. Satisfacció

Aquesta variable es relaciona directament amb l'experiència d'usuari, se centra en com la persona se sent quan utilitza les categories o el cercador de Viquipèdia. L'anàlisi es focalitza en la satisfacció de les persones usuàries (lectores i editores) en diversos aspectes clau de la

Viquipèdia. S'ha mesurat la satisfacció mitjançant una puntuació de 0 a 5, on 0 representa "insatisfactori" i 5 indica "totalment satisfactori". A continuació, es comenten els resultats més rellevants de la taula 17.

Taula 17. Satisfacció amb sistemes de cerca i navegació de la Viquipèdia per cada una de les tasques plantejades

Satisfacció	Tasca 1 [Navegació autònoma]	Tasca 2 [Cercador general]	Tasca 3 [Cercador de gènere]	Tasca 4 [Cerca interna]	Tasca 5 [Cercador d'intersecció]	Tasques 6-8 [Categories en arbre]
Mitjana	3,5	3	2,5	2	2	2,9
Mitjana (editors)	3,9	3,6	2,4	2,3	2,3	2,8
Mitjana (lectors)	2,9	2,8	2,1	1,7	1,8	1

Tasca 1 “Navegació autònoma”: La mitjana de satisfacció en aquest aspecte és de 3,55, indicant que la navegació autònoma és percebuda de manera positiva per persones usuàries i editores. Les editores mostren una satisfacció lleugerament superior, amb una mitjana de 3,9, en comparació amb la mitjana dels lectors de 2,9.

Tasca 2 “Cercador general”: La mitjana de satisfacció en aquest aspecte és de 3, pels dos grups que mostren un nivell similar de satisfacció en l'ús del cercador general.

Tasca 3 “Cercador de gènere”: La mitjana de satisfacció en aquest aspecte és de 2,5, les persones lectors mostren una satisfacció moderada en aquest aspecte, i en canvi les editores presenten una puntuació lleugerament superior amb una mitjana de 2,44. Aquest cercador és el que es va implementar després del darrer procés de deliberació dins la comunitat viquipedista per donar accés d'alguna manera a les diferents identitats de gènere, ja que es va prohibir l'ús de la categoria “dones” i “persones no-binàries”.

Tasca 4 “Cerca interna”: La mitjana de satisfacció en aquest aspecte és de 1,9. Tant les editores com les lectors mostren una satisfacció relativament baixa en l'ús de la cerca interna.

Tasca 5 “Cercador d'intersecció”: La mitjana de satisfacció en aquest aspecte és d'1,95. Tant les usuàries com les editores mostren un nivell baix de satisfacció en l'ús del cercador d'intersecció.

Tasques 6-8 “Categories en arbre”: La mitjana de satisfacció en aquest aspecte és de 2,9. Si bé les persones lectors i editores mostren una satisfacció moderada en aquest aspecte, les editores tenen una puntuació lleugerament superior amb una mitjana de 2,8.

Les eines considerades més satisfactòries són l'arbre de categories i el cercador de gènere. També s'ha assenyalat la importància d'una cerca general ben configurada i la navegació per categories. La navegació autònoma i el cercador general reben una satisfacció moderada per part dels usuaris i editors, la cerca interna i el cercador d'intersecció generen un nivell de satisfacció més baix.

En les preguntes obertes dels tests amb persones usuàries s'ha destacat que la navegació i cerca a la Viquipèdia pot ser particularment difícil si la persona usuària no té coneixements detallats sobre l'estructura i el funcionament del lloc. Aquesta manca d'intuïció en la cerca pot generar frustració i dificultats per trobar la informació desitjada de manera eficient. En conseqüència, les persones usuàries han suggerit que si no es visibilitzen les categories és molt complicat accedir a la informació de Viquipèdia.

En general, les persones usuàries han fet evident que la interfície és confusa i poc intuïtiva, i que s'han trobat amb dificultats per completar tasques de cerca i recuperació d'informació. La recuperació d'informació a través de la Viquipèdia és percebuda com a limitada en comparació amb la utilització de cercadors externs. La cerca a la Viquipèdia requereix un coneixement detallat de l'estructura i el funcionament del lloc, incloent la visibilització de les categories. Aquests factors poden contribuir a una experiència frustrant per a les persones usuàries especialment per les persones que no són viquipedistes.

5.2.2. Avaluació heurística

L'avaluació heurística de l'esquema de categories de Viquipèdia s'ha realitzat a partir de la tècnica de la inspecció d'estàndards en què un expert en usabilitat analitza si la interfície segueix les especificacions acordades i els estàndards definits en l'àmbit internacional. En el cas que ens ocupa s'han generat un conjunt d'indicadors identificats, especialment, a partir de fonts normatives. Aquestes fonts ens proporcionen també mètodes d'obtenció d'evidències per a cada indicador, les mètriques aplicades i, quan és possible, valors òptims. En l'apartat "2.5. Proposta d'avaluació heurística de l'esquema de categories de Viquipèdia" es detallen aquests components de l'avaluació heurística. A continuació es presenten els resultats obtinguts.

5.2.2.1. Avaluabletat

L'esquema de categories de Viquipèdia és avaluable perquè els creadors de categories disposen de diferents tipus d'eines consensuades per a la seva pràctica. En destaquem les següents.

- Directriu sobre Categorització⁵⁰ resultant del procés de discussió i presa de decisions propi de l'enciclopèdia⁵¹.
- Ajudes per als creadors de categories: Ajuda:Categoria⁵² i secció "Llibre d'estil sobre categorització" en la directriu sobre Categorització⁵³.
- Plantilles per als creadors de categories: Categoria:Plantilles de manteniment per a categories⁵⁴.
- Hi ha un control sobre les pàgines que no contenen categories, per als treballs de manteniment.

Hem contrastat, informalment, el grau de coneixement d'aquestes eines amb alguns administradors individuals de la comunitat viquipedista i ens han traslladat el seu desconeixement sobre elles.

⁵⁰ Wikimèdia. (2022). Viquipèdia:Categorització. En *Viquipèdia, l'enciclopèdia lliure*.

<https://ca.wikipedia.org/w/index.php?title=Viquip%C3%A8dia: Categoritzaci%C3%B3&oldid=3024068>. És una directriu.

⁵¹ Vegeu apartat "4.3. Anàlisi de la creació i processos de presa de decisions dels sistemes d'organització del coneixement de Viquipèdia: categories" per a entendre el marc de deliberació de Viquipèdia.

⁵² Wikimedia. (2018). Ajuda: Categoria. En *Viquipèdia, l'enciclopèdia lliure*.

<https://ca.wikipedia.org/w/index.php?title=Ajuda: Categoria&oldid=20513864>. És una ajuda.

⁵³ Wikimèdia. (2022). Viquipèdia:Categorització. En *Viquipèdia, l'enciclopèdia lliure*.

<https://ca.wikipedia.org/w/index.php?title=Viquip%C3%A8dia: Categoritzaci%C3%B3&oldid=3024068>. És una directriu.

⁵⁴ Wikimedia. (2015). Categoria: Plantilles de manteniment per a categories. En *Viquipèdia, l'enciclopèdia lliure*.

<https://ca.wikipedia.org/w/index.php?title=Categoria: Plantilles de manteniment per a categories&oldid=16026819>.

5.2.2.2. Reusabilitat

Cada categoria té una única instància i un únic identificador amb independència de les diferents localitzacions que pugui tenir en les jerarquies de l'esquema.

Les categories de l'esquema de Wikipedia estan disponibles per al bolcat (database dumps) en tres formats d'intercanvi de dades: sql, json i xml. Aquests formats no proporcionen informació semàntica sobre els conceptes i les relacions entre els conceptes de l'esquema de categorització del conjunt de Wikipedia, com podria aportar el model de dades Simple Knowledge Organization System (SKOS). D'aquesta forma, resulten molt limitades les possibilitats de reutilització de les categories de Wikipedia en altres conjunts de dades, o en sistemes de recuperació d'informació.

5.2.2.3. Estabilitat

Les magnituds més destacables relatives a l'estabilitat durant el període 2004-2022 són les següents:

- Nombre mínim de noves categories: 866 (2005).
- Nombre màxim de noves categories: 8.616 (2021).
- Mitjana de noves categories per any: 1.769,57.
- Mediana de noves categories: 5.621 (2014).

La taxa de creixement anual es va mantenir alta durant el període 2004-2007, i, a partir del 2008 va experimentar un descens pronunciat fins al 2012. A partir d'aquest any, la taxa mostra com l'increment es va reduint progressivament, i la taxa s'estabilitza, fins al 2020 on va experimentar un increment molt destacable. Aquest increment pot estar vinculat a un dels efectes de la pandèmia del Covid-19: la disponibilitat més gran de temps per a la contribució en la Viquipèdia.

Figura 1. Indicador Estabilitat

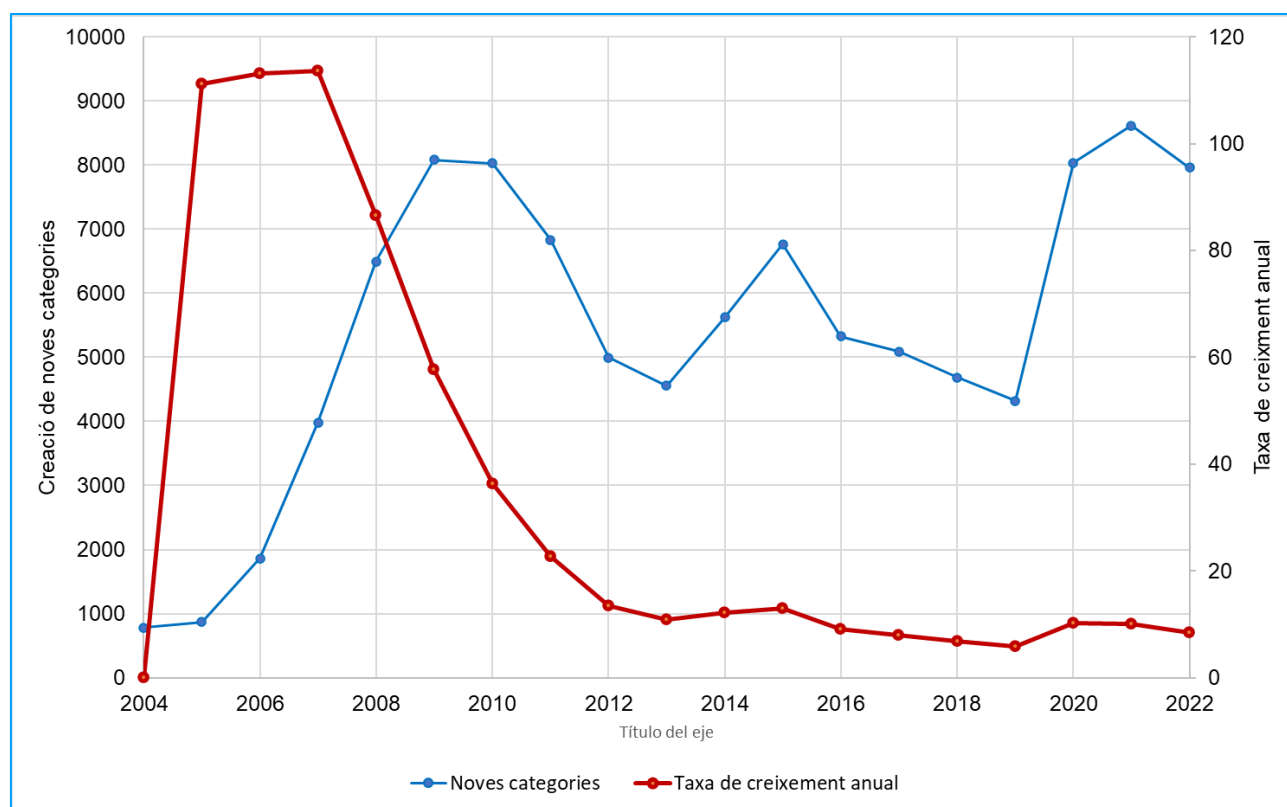


Figura 2. Indicador Estabilitat. Categories i acumulat

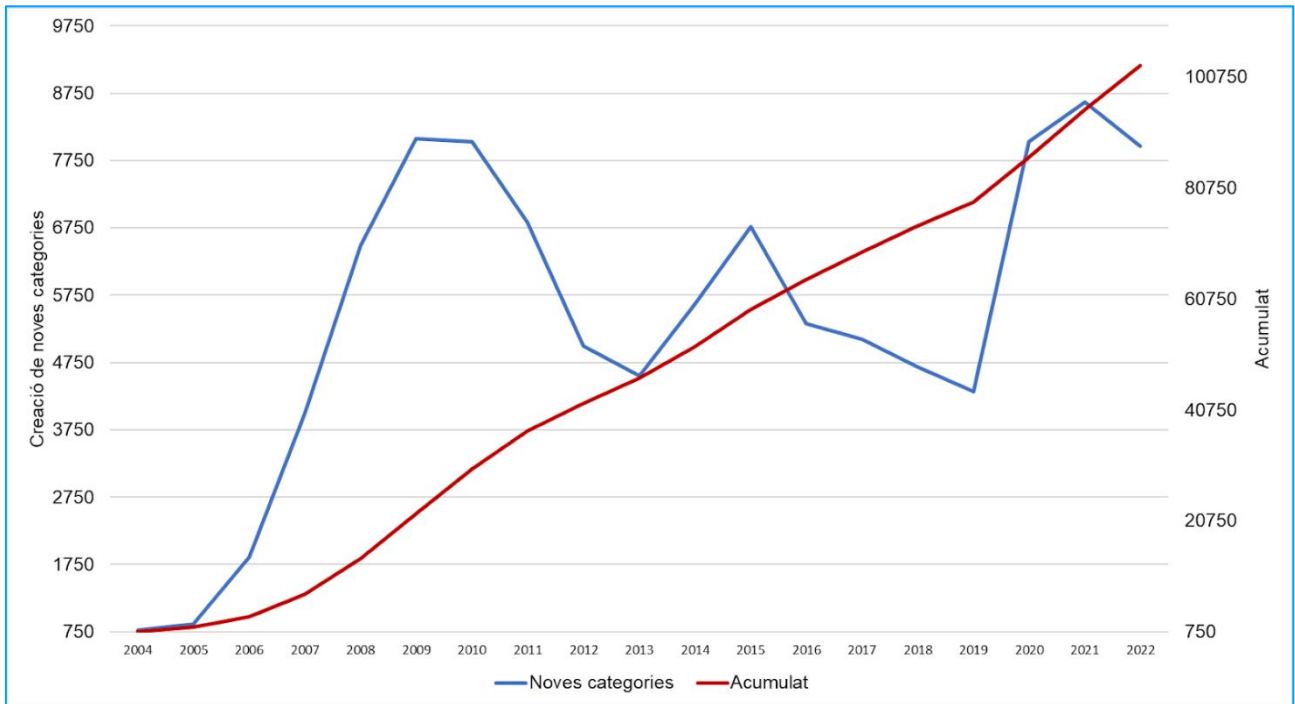
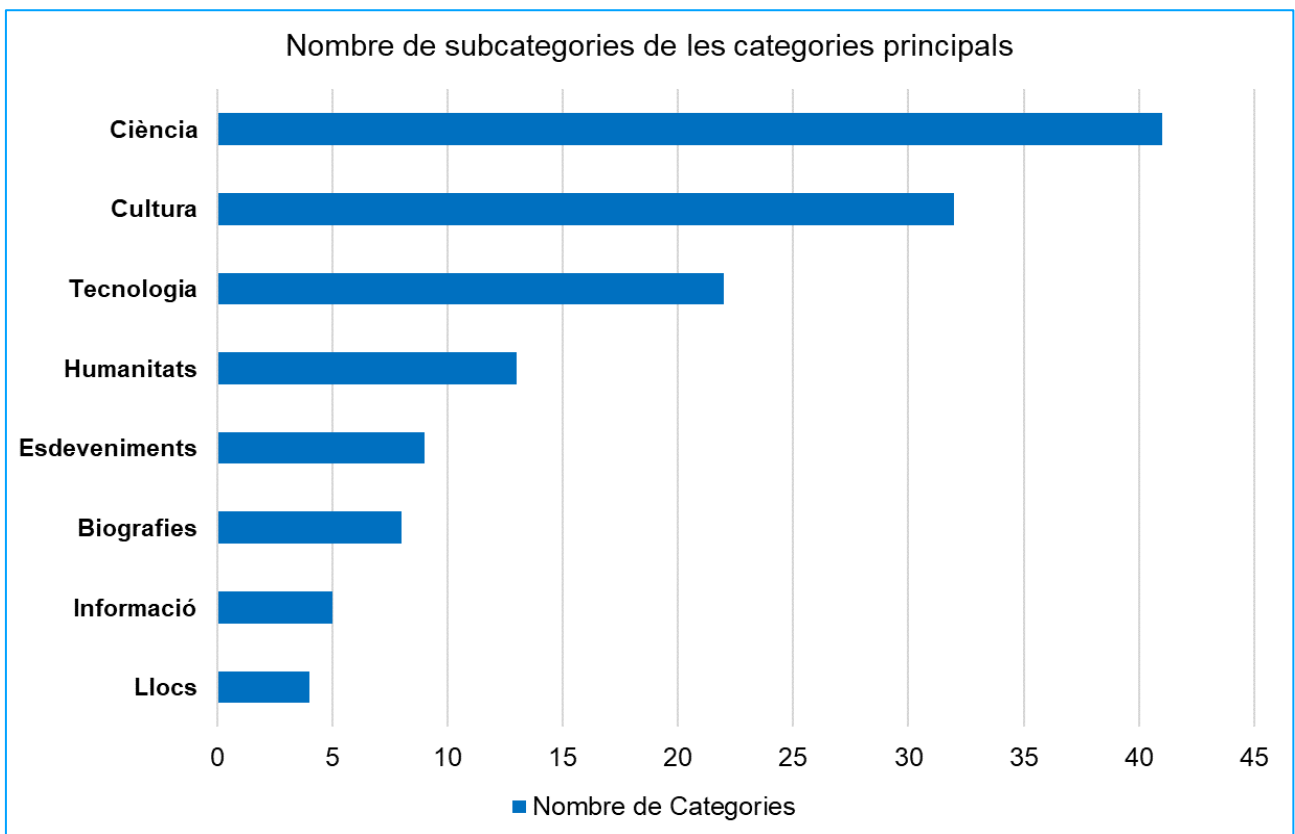


Figura 3. Indicador: Estabilitat. Categories i Acumulat



5.2.2.4. Nombre de categories (conceptes)

El 31 de desembre de 2022, l'esquema de categories de Viquipèdia incorporava 102.159 categories. Podem comparar aquesta mida amb el d'altres SOC de característiques similars:

estan formats per conceptes precoordïnats, i la seva orientació és representar el coneixement enciclopèdic.

La *Llista d'encapçalament de matèries en català* (LEMAC) conté 112.200 encapçalaments, considerant tant els acceptats com els no acceptats. El punt de partida és la traducció de la versió en castellà de *Library of Congress Subject Headings* (LCSH), que es va publicar en edició preliminar pel Servei de Biblioteques de la Generalitat de Catalunya l'any 1988. És difícil realitzar comparacions pel que fa a la taxa de creixement entre l'esquema de categorització de Viquipèdia i LEMAC, perquè el primer tot just es troba en els seus primers vint anys d'existència, mentre que LEMAC assoleix els 35 anys, i molts més si tenim en compte el seu origen. No obstant això, els clars indicis d'un creixement més ràpid de l'esquema de categorització de Viquipèdia es constata en les següents mostres.

- L'any 2009, el creixement de LEMAC va ser de 2.663 encapçalaments nous, enfront de 8.080 categories noves en Viquipèdia.
- L'any 2021, el creixement de LEMAC va ser de 1.219 encapçalaments nous, enfront de 8.616 categories noves en Viquipèdia.

No és comparable la magnitud de categories de Viquipèdia, i la seva taxa de creixement, amb els de SOC aplicats a enciclopèdies disponibles en l'entorn digital. L'arbre de coneixement de la *Gran Enciclopèdia Catalana* (GEC) conté 425 categories, i les categories de la *Encyclopedia Britannica* són 123. En ambdós casos, la jerarquia del SOC es limita a dos nivells.

5.2.2.5. Nombre de relacions semàntiques

No tenim accés a les dades de totes les relacions de jerarquia entre supercategories i subcategories. Sí disposem d'un còmput parcial que inclou els tres primers nivells de l'esquema de categorització; és a dir, relacions entre les categories temàtiques principals i el segon nivell, i entre les categories de segon nivell i les categories de tercer nivell. En aquest àmbit, el resultat és de 1122 relacions de jerarquia.

5.2.2.6. Índex d'enriquiment o granularitat

En els tres primers nivells de l'esquema de categories hi ha 143 categories i 1122 relacions de jerarquia. A partir d'aquestes magnituds, la mitjana corresponent a l'índex d'enriquiment o granularitat és de 7,8461. Supera, amb escreix, el valor òptim situat entre 2 i 5 (vegeu Taula 9).

5.2.2.7. Grau de precoordïnació

En el conjunt de l'esquema de categories de Viquipèdia, la mitjana de paraules que conformen les designacions de les categories és de 3,6766. Supera el valor màxim situat entre 1,5-2 que estableixen els estàndards i recomanacions dels especialistes (vegeu Taula 9). Altres dades que reflecteixen la vulneració d'aquest valor òptim són les següents:

- Mediana de paraules que conformen les designacions de les categories: 3 paraules.
- Nombre màxim de paraules que conformen les designacions de categories: 18 paraules en *Resolucions del Consell de Seguretat de les Nacions Unides sobre el Tribunal Penal Internacional per a l'antiga Iugoslàvia*.

5.2.2.8. Nombre de nivells en la jerarquia o profunditat

Per a l'avaluació d'aquest indicador, hem revisat totes les cadenes jeràrquiques de la categoria temàtica principal *Biografies*. El resultat ha estat que, en tots els casos, hi ha més de cinc nivells. En conseqüència, no es compleix el valor màxim que estableixen els estàndards i recomanacions dels especialistes (vegeu Taula 9).

5.2.2.9. Nombre de categories en un mateix nivell de la jerarquia o amplitud

Per a l'avaluació d'aquest indicador, hem revisat els dos primers nivells de subcategorització posteriors a les vuit categories temàtiques principals. En conjunt, aquesta secció conté 144 categories, de les quals 130 inclouen subcategories, mentre que les 14 restants enllacen directament amb pàgines de Viquipèdia.

Les magnituds d'aquestes 130 categories són les següents.

- Hi ha 17 que se subcategoritzen amb una sola subcategoria. Representen el 13,07% de la secció avaluada, i, de fet, constitueix la seva moda. Aquests casos vulneren el mínim de dues subcategories que estableixen els estàndards i recomanacions dels especialistes.
- Hi ha 35 categories que se subcategoritzen amb més de 12 subcategories. Representen el 26,92% de la secció avaluada. Aquests casos vulneren el màxim de dotze subcategories que estableixen els estàndards i recomanacions dels especialistes. La màxima superació del límit es produeix en la categoria *Religió*, que inclou 43 subcategories.

Sumant les infraccions del mínim i del màxim de subcategories, tenim que el 40% de les categories avaluades vulneren els valors òptims de l'*amplitud* (52 de 130 categories).

5.3. Anàlisi del procés de presa de decisions

Per crear una categoria a Viquipèdia, qualsevol persona ho pot fer. Algunes estableixen que cal tenir un mínim de 5 articles, algunes 2, i a partir de tenir algun contingut concret, ja es pot etiquetar en una categoria per afavorir la seva visibilitat i conseqüent accés a la informació. Si algun membre de la comunitat viquipedista s'adona de la creació d'una nova categoria, o revisa les anteriors, pot decidir engegar un procés per esborrar-la. D'una banda, hi ha persones amb privilegis per fer-ho (rol dels administradors), i de l'altra es pot promoure processos de presa de decisió consensuada a través de negociacions (deliberació) o de votacions per formar opinió i decidir si cal esborrar-la.

En el cas de la Viquipèdia catalana, i el cas de les etiquetes per gènere, s'han creat en diferents ocasions etiquetes sobre dones amb una professió o condició comuna, i sempre s'han esborrat. Un exemple recent seria l'etiqueta “dones” o “LGTBI” que van ser esborrades com a categories després que un viquipedista les creés. El cas de “dones” després d'un llarg procés de deliberació.

El sistema de presa de decisions a Wikipedia depèn de cada comunitat. A la Viquipèdia (Wikipedia catalana), la comunitat pot decidir sobre diversos temes: crear algun article en especial, les responsabilitats dels usuaris, la supressió de pàgines, disseny de la plana web, elaboració de plantilles, repartiment de tasques, etc. Les negociacions es donen íntegrament dins de l'espai de Viquipèdia, ja sigui en els espais de Discussió de cada pàgina o espais de Deliberació que es creïn o bé en espais de Votació per acabar de tancar un tema⁵⁵. I com a norma general, la Viquipèdia no funciona per majoria sinó que és per consens, després que es puguin haver donat llargs processos de negociacions en el que es busca anar formant una opinió conjunta i es vagin aportant idees.

En el sistema de presa de decisions de les categories de la Viquipèdia s'hi donen els processos de deliberació i els de votació. En els processos de deliberació hi pot participar qualsevol persona usuària registrada a Viquipèdia, però en les votacions no hi pot participar tothom. Per exemple

⁵⁵ https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Sistema_de_presa_de_decisions

en les pàgines en què es decideix què és i què no és coneixement a la Viquipèdia (Esborrar pàgines/Propostes⁵⁶) només hi poden votar, o només són vàlids els vots, de les persones que fa més de 30 dies que són viquipedistes, que han fet més de 100 contribucions a l'espai principal (articles) en els darrers tres mesos.

En el cas que ens ocupa en aquest estudi, que és la decisió que es va adoptar de no usar les categories de “dones” i “persones no-binàries”, hi va haver tres processos de deliberació a Viquipèdia, on les persones editores van expressar els seus arguments a través de comentaris per tal que la comunitat pugui generar opinió i es pogués prendre decisions. En els processos de negociació, qualsevol persona usuària de Viquipèdia hi va poder participar aportant arguments perquè justament és un procés per conèixer l'opinió de la comunitat i buscar solucions conjuntes. En el procés de votació, en canvi, no tothom hi podia participar i qui ho va fer sense complir els requeriments se li va anul·lar el vot.

Les categories de gènere, a banda dels tres processos de negociació (o deliberació en terminologia viquipedista) també ha tingut un procés de votació en què l'opinió majoritària va ser en contra de les categories de gènere. En aquests processos de votació no hi pot participar tota la comunitat viquipedista, només poden emetre vot les persones usuàries amb 500 edicions a les pàgines principals (articles) en els darrers 3 mesos⁵⁷.

5.3.1. Anàlisi del procés de presa de deliberació

A l'apartat de Metodologia 2.6 s'ha explicat la manera com s'ha treballat l'anàlisi dels arguments del procés de deliberació, i en aquest apartat presentarem els resultats de l'anàlisi.

En total a les tres deliberacions hi van participar 23 persones editores (12 a la primera deliberació, 15 i 3 respectivament). Majoritàriament homes 15, 6 dones i 2 de gènere neutre. La representativitat dels resultats de les deliberacions és molt baixa, ja que el total de persones que van debatre la qüestió representa el 0,005% de les persones registrades a Viquipèdia com a persones usuàries. La participació va comptar amb 233 comentaris (99 a la primera, 128 a la segona i 6 a l'última). Els homes van fer el 83,7% dels comentaris (195) i les dones el 34% (38).

A la taula 18 es presenten els posicionaments dels viquipedistes sobre les categories de “dones” i “persones no-binàries” expressats en les tres deliberacions (D1, D2 i D3) segons gènere home/dona o no especificat. Així, en total, es van obtenir 82 comentaris posicionats contra les categories d'identitat de gènere (61,4%) i 42 a favor (38,5%). És important tenir present que la posició majoritària va ser “en contra però amb propostes” amb 107 posicionaments. La deliberació més participada va ser la segona sobre la categoria “dones” (103 posicionaments), especialment amb perspectives “en contra però amb propostes” (87) que són els posicionaments que rebutgen les propostes, però busquen solucions consensuades. Les dones són les que es van posicionar més en contra.

Els posicionaments d'usuaris masculins destaquen per proporcionar més diversitat, tot i que cal tenir en compte que hi participen més (150 posicionaments en comparació dels 43 de les dones). Tot i que en general els homes es postulen en contra o en contra amb propostes la diversitat d'identitat de gènere (39,1% neutral -en contra amb propostes-, 17,9% en contra i 15,5% a favor).

A l'última deliberació sobre “persones no-binàries” la discussió va ser molt breu, d'unes hores de durada, i els resultats van ser majoritàriament a favor de la creació de la categoria (2 contra 1).

⁵⁶ https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Esborrar_p%C3%A0gines/Propostes

⁵⁷ https://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Sistema_de_presa_de_decisions

Taula 18. Posicionament d'arguments en les deliberacions de l'acceptació o rebuig de les categories “dones” (D1 i D2) i “persones no-binàries” (D3)

Deliberacions	Període de negociació i vot	Orientació del vot			Participació	
		En contra però amb propostes	En contra	A favor	Nombre de comentaris	Nombre de participants
D1 “categoria: dones”	2014 (27 dies)	30	56	27	99	12
D2 “categoria: dones”	2015-17 (2 anys)	57	16	15	128	15
D3 “persones no-binàries”	2022 (1 dia)	0	2	4	6	3

5.3.2. Anàlisi dels arguments de les deliberacions

Els arguments de les deliberacions s'han analitzat a partir de les codificacions de la tipologia d'arguments i del posicionament dels mateixos cap a l'acceptació o el rebuig de les categories d'identitat de gènere proposades (“dones” i “persones no binàries”) tal com s'ha explicat a l'apartat 2.6 de la Metodologia.

El tipus d'argumentació més utilitzada pels participants en contra de l'acceptació de les categories de gènere es basa en criteris o opinions personals (argumentació “de valors”, 31% dels arguments) o que fan referència a criteris o opinions personals que ja havia comentat una altra persona en la mateixa deliberació (“sense raonament” 18%), seguits per arguments del tipus “de gènere” (9%) o de “desconeixement” (8%). (Vegeu la taula 10 per recordar la descripció de cada tipologia d'argument). A escala qualitativa, una participant en la discussió va expressar que “ara ja s'està bé perquè és embolic posar-se amb el tema de la gent que no s'identifica amb el seu gènere de naixement, no tenim per què estar al cas de casos particulars, a més ara hi ha tants gèneres diferents que segrega. Som tots iguals i tots anem al mateix sac”. Un altre participant que també compartia l'opinió contrària a categoritzar per gènere, li semblava que seria “sexista per a les dones perquè hi ha criteris molt més rellevants” per categoritzar les persones.

En canvi, els arguments que es mostren favorables a la visibilització de les identitats de gènere a la Viquipèdia es basen en argumentacions de tipus “raonament pràctic” per intentar aconseguir el consens (22%), de “precedents o jurisprudència” que se sustenta a partir de fer referències a altres deliberacions ocorregudes o citar d'altres consensos aconseguits en altres comunitats (19%), arguments del tipus “evidència cap a la hipòtesi” (16%) en què s'aporten referències o dades, o del tipus “analogia” a casos similars (11%). Entre els comentaris compartits a la deliberació es feia referència a deliberacions anteriors, o a articles. En un cas, una de les participants basant-se en els coneixements de la seva professió (“arguments de posició de coneixement”) compartia que les categories per gènere no segregarien sinó que afavoririen l'accés a uns continguts que ara estan menys accessibles. I un altre participant va destacar com a analogia que ja existeix a la Viquipèdia categories que utilitzen la distinció per gènere com el cas de la categoria “dones barbudes”.

Pel que fa al tercer posicionament que apareixia a les deliberacions en contra de les categories, però amb opcions de millorar la visibilitat de les categories de gènere (“rebuig amb propostes”),

bàsicament aportaven els tipus d'argument de "raonament pràctic" per intentar aconseguir el consens (32 arguments). És a dir que s'oposaven a les categories, però promovien el trobar solucions per implementar i així aconseguir un consens. Entre les propostes que es van presentar en les negociacions va sortir l'ús de plantilles amb notes a peu de pàgina o crear enllaços com a "Ciència en femení".

És interessant destacar que malgrat que hi van participar molt poques dones (6), totes es van posicionar al rebuig de les categories de "dones" però a l'opció de trobar millores per visibilitzar-les.

5.3.3. Anàlisi del procés de votació

En el procés de votació de 2018 van votar 28 persones, però 7 es van anul·lar per no complir els requisits d'antiguitat i llindar de producció de continguts establerts per la Viquipèdia. Entre els votants vàlids hi va haver només 4 dones, 16 homes i una persona que no especificà el gènere.

El resultat de la proposta va ser el de rebutjar la proposta de visibilitzar i donar accés a continguts segregats amb les categories relacionades amb la identitat de gènere (categoria de "dones") amb un total de 18 vots. D'aquests, 14 van votar que no a la categoria per gènere, però que calia incorporar millores tècniques per facilitar-ne la recuperació dels continguts. I només 3 persones. Es van poder emetre diferents tipus de vots, un era un sí rotund a la categoria de gèneres (3 vots) i després hi havia diferents tipus de vots negatius a la categoria. Un era el no rotund (4 vots), i d'altre rebutjaven les categories, però proposaven opcions de millora (14 vots) (vegeu Taula 19). Entre aquests darrers tipus de vots desfavorables amb opcions de millora hi havia propostes de solucions com ara crear un cercador que s'implementés només en categories de professions feminitzades, crear categories per a totes les identitats de gènere possibles, o només quan la identitat es pugui fonamentar en fonts fiables. Finalment, l'opció més votada va ser no a la creació de la categoria "dones" però amb l'opció de crear un cercador (una plantilla de consultes).

Taula 19. Votació sobre la creació de la categoria de "dones" a la Viquipèdia catalana

Tipus de vot proposat	Significat	Nombre de vots	Sentit del vot
"Categoria per gènere (CxG) sense restriccions"	'sí' a la categoria de "dones"	3	FAVORABLE
"Categoria per gènere (CxG)"	'no' a la categoria de "dones"	4	DESFAVORABLE
"No fer Categoria per gènere (CxG), però utilitzar la plantilla de cerca de gènere"	'no' a la categoria de "dones" però 'sí' a una proposta de crear un cercador específic per gènere a través de Wikidata	14	DESFAVORABLE

5.4. Conclusions de la presència o absència de biaix de gènere

A partir del marc teòric revisat, i de l'anàlisi i avaluació del sistema d'organització de coneixement de Viquipèdia com a sistema de categorització, arribem a les següents conclusions.

- La revisió de la literatura acadèmica no discrepa en cap cas que **Wikipedia pateix d'un biaix de gènere persistent, detectat des de 2011, tant en els continguts com en la diversitat de les persones voluntàries que editen els continguts en qualsevol llengua. També hi ha un biaix de gènere en el sistema d'organització de continguts de Viquipèdia com aquest estudi demostra.**
- **El caràcter enciclopèdic de Viquipèdia ha primat la cosmovisió de les persones que indexen i classifiquen les categories en detriment de les necessitats de les persones usuàries.** Aquest aspecte ha determinat que les categories s'han dissenyat per agrupar pàgines sota conceptes i per facilitar la consulta per necessitats concretes; sobretot les necessitats de les mateixes persones editores. I, en canvi, no s'ha dissenyat tant per complir la funció de recuperació de la informació i satisfer les necessitats informatives de les persones usuàries.
- **El sistema de navegació i cerca de Viquipèdia presenta marge de millora des del punt de vista de la usabilitat, l'experiència d'usuari i dels estàndards:**
 - El cercador per gènere es valora positivament, però no es fa visible ni trobable, de forma fàcil i ràpida, a les persones usuàries que el necessiten per a cercar informació. Només un 30% dels participants en els experiments coneixien la seva existència.
 - Les categories que organitzen els articles no són advertides per les persones lectores i, per tant, no les utilitzen en les seves estratègies de navegació.
 - Les persones usuàries prefereixen utilitzar cercadors externs per a recuperar informació abans que els sistemes de navegació i cerca que proporciona Viquipèdia.
 - En SOC de la Viquipèdia no compleix molts dels indicadors de qualitat establerts en els estàndards i especificacions dels especialistes. Per exemple:
 - El nombre de nivells de les jerarquies (profunditat) vulnera, sovint, el valor màxim recomanat. Aquest valor màxim s'ha situat en 12 nivells.
 - El nombre de subcategories d'una categoria (amplada) vulnera, sovint, els valors mínim i màxims recomanats. Aquests valors s'han situats en 2 i 5 subcategories, respectivament.
- L'avaluació de les **categories** de Viquipèdia a partir dels estàndards dels sistemes d'organització del coneixement permet concloure que **hi ha molt marge de millora, no només en els aspectes d'identitat de gènere, sinó en el sistema d'organització del coneixement global per a facilitar la cerca i recuperació a les persones usuàries.**
- En les negociacions de la comunitat viquipedista en què es van rebutjar les categories per gènere, es detecta un component social i cultural **en els arguments contraris a les categories de gènere, el qual s'allunya del coneixement científic-tècnic ja acceptat en estàndards internacionals referents en els sistemes d'organització del coneixement.**
- Des del punt de vista de la **perspectiva de gènere i l'anàlisi interseccional**, es detecta que **no hi ha un criteri objectivable en el rebuig de les identitats de gènere (“dones” o “persones no-binàries”) com a criteri de categorització a Viquipèdia**, ja que actualment s'accepten categories com ara “origen”, “religió” o “estatus”. Al mateix temps, es detecten inconsistències com l'ús i acceptació de categories de primer nivell en femení com “llevadores” o “dones barbudes”.
- Tenint en compte que el sexe és un element biològic i el gènere és un element social, referenciables amb fonts d'informació externes i contrastades, **no hi ha cap raó des del**

punt de vista dels fonaments i principis de la Viquipèdia que impedeixi que l'element identitari del gènere esdevingui un criteri de classificació.

- **Existeixen referents que incorporen el gènere com a criteri d'organització del coneixement per satisfer les necessitats de les persones usuàries. Així per exemple, les biblioteques que tradicionalment han garantit l'accés universal al coneixement a través de sistemes que l'organitzen, han incorporat el criteri de gènere des de fa moltes dècades.**

La totalitat de les Wikipedies, amb la sola excepció de la catalana i la italiana, utilitzen amb normalitat les categories d'identitats de gènere per facilitar la cerca i recuperació dels continguts que contenen.

6. Diagnosi de Wikidata

En aquesta secció s'analitza el sistema d'organització del coneixement per ontologies del *dataset* Wikidata que dona suport i empodera a l'enciclopèdia digital i col·laborativa Wikipedia.

S'avalua el sistema amb dos mètodes, test amb persones usuàries (lectores i editores) i test heurístic elaborat a partir de l'observació de persones expertes (avaluació heurística) amb inspecció d'estàndards.

Finalment, s'aporten les conclusions de la presència o absència del biaix de gènere a partir d'aquesta diagnosi.

6.1. Anàlisi del sistema d'organització del coneixement per ontologies

Wikidata és una base de coneixement a gran escala, Wikidata (Vrandečić & Krötzsch, 2014). És una base de dades gratuïta, col·laborativa (a) i multilingüe, que inclou gairebé 300 llengües (Kaffee et al., 2017), que serveix com a base de dades secundària (b) i recopila dades estructurades per donar suport a altres projectes del Wikiverse com ara la Wikipedia (enciclopèdia), Wikimedia Commons (repositori d'imatges, àudios i audiovisuals), així com a qualsevol persona del món.

a. És col·laborativa perquè les dades són introduïdes i mantingudes per persones editores, que poden ser humans o robots automatitzats (bots)⁵⁸, i es poden manifestar mitjançant autenticació o de forma anònima. Els editors humans decideixen les regles de creació de gestió de continguts. Fins i tot permet l'edició col·laborativa del seu model de dades per una àmplia comunitat d'usuaris (Piscopo et al., 2017).

b. És secundària perquè no només registra descripcions sobre objectes i conceptes, sinó també les fonts que avalen aquestes descripcions, i les connexions amb bases de dades externes que les complementen.

6.1.1. Model de dades

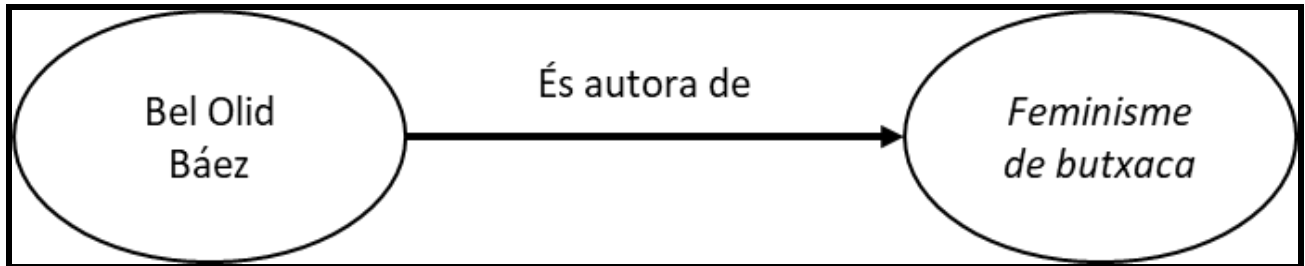
Wikidata s'estructura a través d'un model de dades que en facilita la recuperació de les seves pròpies dades. El model que ha optat Wikidata és el model de dades orientat a grafs, i, més específicament, grafs dirigits. La unitat mínima d'un graf dirigit són dos nodes vinculats per una aresta.

⁵⁸ <https://wikidata.org/wiki/Wikidata:Bots>

- Un dels nodes és un objecte o concepte del món sobre el qual volem expressar una dada.
- L'aresta expressa un atribut del primer node.
- El segon node és el valor de l'atribut que l'aresta descriu sobre el primer node.

Podem observar una representació gràfica d'una unitat mínima d'un graf en la Imatge 10.

Imatge 10. Representació gràfica d'una unitat mínima d'un graf



Cada un dels dos nodes pot vincular-se a altres nodes mitjançant arestes, és a dir propietats. Es tracta, doncs, d'una estructura molt simple que permet l'expansió sense un límit previst en l'origen, i, en un context de dades obertes, sense els límits d'una única corporació.

El model de dades orientats a grafos ha estat objecte de diferents estandarditzacions. Una d'elles, és Resource Description Framework (RDF), portada a terme pel W3C en el marc del seu projecte de web semàntic⁵⁹ que vol dotar de contingut semàntic a les dades del Web, i aprofitar aquest contingut per a millorar l'eficàcia i l'eficiència dels processos de navegació i cerca. El model de dades RDF designa l'estructura formada per dos nodes vinculats per una aresta *declaració o tripleta RDF*. El node que és descrit constitueix el subjecte de la declaració; l'aresta constitueix el predicat (o propietat de la declaració), i el node que expressa el valor sobre el subjecte és l'objecte de la declaració.

RDF exigeix que el subjecte i la propietat siguin ocupades per recursos identificats amb un Internationalized Resource Identifier (IRI). Un IRI és una cadena Unicode per identificar nodes i vores d'una manera inequívoca. Els IRI són versions internacionalitzades dels URI que són generalitzacions d'URL. L'ús d'IRI ajuda a evitar conflictes de noms i promou la denominació distribuïda, és a dir, per evitar requerir una autoritat de denominació centralitzada. A més, l'IRI pot ser resoluble i accessible a través del protocol HTTP, de forma que el recurs, i les dades sobre el recurs, són accessible de forma inequívoca a través del Web.

Pel que fa a l'objecte de la declaració, el model RDF permet que sigui ocupat o bé per un recurs identificat amb un IRI (igual com succeeix amb el subjecte i la propietat), o per un literal. El primer presenta les mateixes característiques que hem vist en referir-nos al subjecte i propietat de les declaracions. En canvi, el literal és un valor concret utilitzat per a representar tipus de dades com cadenes de caràcters, nombres, dates, etc.

El model de dades RDF, i el projecte de web semàntic en el seu conjunt, van estar en l'origen de Wikidata. Aquest, però, va renunciar a seguir completament els requeriments del model de dades RDF. Això no obstant, la interoperabilitat entre Wikidata i conjunts de dades RDF està perfectament assegurada. Veiem, a continuació, les especificitats del model de dades a Wikidata⁶⁰.

⁵⁹ https://w3.org/2001/sw/wiki/Main_Page


⁶⁰ Més detalls sobre el model de dades de Wikidata es poden consultar en Wikimedia. (2023). *Wikibase/DataModel*. <https://mediawiki.org/wiki/Wikibase/DataModel>

En la base de dades de Wikidata, el recurs en posició de subjecte és designat element (*item* en anglès). Per a cada element, es proporcionen diverses descripcions en forma de declaracions o, si ho veiem des de la perspectiva dels grafs, diverses arestes etiquetades, amb nodes apuntant als elements descrits. El conjunt de les declaracions sobre un element conformen una pàgina a Wikidata, tot i que també estan disponibles en altres formats directament llegibles i consumibles per dispositius informàtics.

Es distingeixen dos tipus de declaracions: informació bàsica sobre l'element per a facilitar la seva identificació, designació i cerca; i declaracions amb contingut més detallat sobre l'element.


Un exemple del primer tipus es pot observar en la Imatge 11, que correspon a la informació bàsica sobre l'element *Bel Olid Báez* (Q16176248)⁶¹.

Imatge 11. Element “Bel Olid Báez Q16176248” a Wikidata



Element [Discussió](#) Mostra [Mostra l'història](#)

Bel Olid Báez (Q16176248)

persona dedicada a la docència, a la traducció i a l'escriptura  [modifica](#)

Isabel Olid Báez | Bel Olid

▼ En més idiomes
[Configura](#)

Llengua	Etiqueta	Descripció	També conegut com a
català	Bel Olid Báez	persona dedicada a la docència, a la traducció i a l'escriptura	Isabel Olid Báez Bel Olid
anglès	Bel Olid	Catalan writer and translator	
anglès americà	Encara no s'ha definit cap etiqueta	Encara no s'ha definit cap descripció	
espanyol	Bel Olid	persona catalana dedicada a la escritura y la traducción	

[Totes les llengües introduïdes](#)

La informació bàsica inclou etiquetes, descripcions breus i àlies (“també conegut com”) en diferents idiomes.

El segon tipus de declaracions, amb contingut més detallat i específic sobre cada element, constitueixen les descripcions més importants a Wikidata i, en conseqüència, meriten una major atenció. Cada una d'aquestes declaracions estan formades per dues parts: una afirmació (*claim* en anglès) que descriu algun aspecte de l'element; i una llista de referències on es pot verificar l'afirmació. Cal dir que la provisió de referències no és una condició sine qua non per a incorporar una afirmació sobre un element.

En la Imatge 12 es reproduïx una afirmació sobre l'element *Bel Olid Báez* (Q16176248).

⁶¹ La identificació d'elements i propietats de Wikidata es realitza mitjançant l'etiqueta en cursiva en català si està disponible (i, en el seu defecte, en castellà o en anglès), seguit de l'identificador únic entre parèntesi. L'identificador únic s'inicia per Q, en el cas dels elements, i per P, en el cas de les propietats.

Imatge 12. Propietat 21 de sexe o gènere per l'element “Bel Olid Báez (Q16176248)” a Wikidata



The screenshot shows the Wikidata interface for the property 'sexe o gènere' (sex or gender) of the entity 'Bel Olid Báez (Q16176248)'. The current value is 'gènere no-binari'. There are two references listed:

URL de la referència	data de consulta
https://twitter.com/BelOlid/status/1491863812311552001	11 feb 2022
https://www.rac1.cat/a-la-carta/detail/465f7da3-6f9c-4e57-9bc9-02b7ff0af81f?program=voste-primer&section=HOUR	11 feb 2022

Buttons for '+ afegiu una referència' and '+ afegiu un valor' are visible at the bottom right.

L'afirmació es completa amb propietat *sexe o gènere*, i l'objecte d'aquesta propietat (el seu valor), que és l'element *gènere no-binari*. La literalitat d'aquesta afirmació és:

“*Bel Olid Báez* té la propietat *sexe o gènere* el valor del qual l'element *gènere no-binari*”.

Aquest element *gènere no-binari* en posició d'objecte de la declaració pot ser el subjecte d'altres declaracions, fora de l'àmbit de descripció de *Bel Olid Báez* (Q16176248).

En relació amb aquesta afirmació, trobem dues referències de fonts on es pot verificar la seva certesa. La literalitat de la primera d'aquestes referències és:

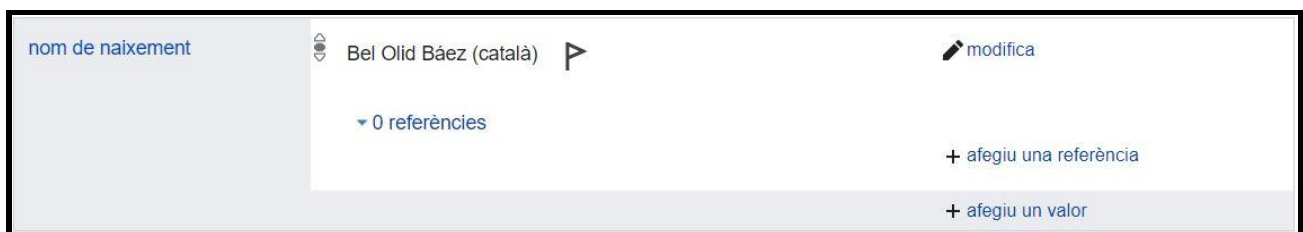
L'afirmació que *Bel Olid Báez* té la propietat *sexe o gènere* el valor de la qual és el *gènere no-binari* està referenciada en la font <https://twitter.com/BelOlid/status/1491863812311552001>. Aquesta font té com a *data de consulta* l'11 feb 2022.

I la literalitat de la segona de les referències és:

L'afirmació que *Bel Olid Báez* té la propietat *sexe o gènere* el valor de la qual és el *gènere no-binari* està referenciada en la font <https://www.rac1.cat/a-la-carta/detail/465f7da3-6f9c-4e57-9bc9-02b7ff0af81f?program=voste-primer§ion=HOUR>. Aquesta font té com a *data de consulta* l'11 feb 2022.

Observem ara l'afirmació sobre l'element *Bel Olid Báez* (Q16176248) que es presenta en la Imatge 13.

Imatge 13. Nom de naixement per l'element “Bel Olid Báez (Q16176248)” a Wikidata



The screenshot shows the Wikidata interface for the property 'nom de naixement' (birth name) of the entity 'Bel Olid Báez (Q16176248)'. The current value is 'Bel Olid Báez (català)'. There are 0 references listed.

URL de la referència	data de consulta
----------------------	------------------

Buttons for '+ afegiu una referència' and '+ afegiu un valor' are visible at the bottom right.

La seva literalitat és:

L'element *Bel Olid Báez* (Q16176248) té la propietat *nom de naixement* el valor del qual és la cadena de caràcters “Bel Olid Báez”. Aquesta cadena de caràcters està expressada en català.

En aquest cas, l’afirmació inclou com a objecte un literal, més concretament, una cadena de caràcters. A diferència de l’element en posició d’objecte que hem vist en l’afirmació anterior, el literal “Bel Olid Báez” no podrà constituir el subjecte d’altres declaracions. La seva funció és, únicament, proporcionar un valor a la propietat *nom de naixement* relativa a l’element *Bel Olid Báez* (Q16176248).

Finalment, observem que, en relació amb l’afirmació de la Figura ZZZ, no hi ha cap referència que la verifiqui (el valor és 0). Tal com succeeix en la Viquipèdia, els editors poden crear afirmacions sense referències. I, igual com en aquella, poden ser millorades per altres editors que sí disposin de referències.

El conjunt de declaracions sobre un mateix element (informació bàsica, afirmacions, referències...) conformen una pàgina web; la pàgina web de l’element. En l’exemple sobre l’element *Bel Olid Báez* (Q16176248), la pàgina web es localitza amb la URL <https://wikidata.org/wiki/Q16176248>, i és accessible mitjançant un navegador, cercable a través d’enginyers de cerca del web, etc. No obstant, més enllà dels límits d’aquesta pàgina, l’entitat *Bel Olid Báez* (Q16176248) s’identifica amb una URI, <http://wikidata.org/entity/Q16176248>, i les declaracions que la descriuen formen part del graf de coneixement de Wikidata, estan accessibles en formats de serialització d’RDF (per exemple, RDF/XML, JSON-LD...), i poden ser consultades i reutilitzades amb llenguatges i mecanismes diferents als del web.

En totes les declaracions que conformen les afirmacions i les referències subjau una ontologia, és a dir, l’especificació d’una conceptualització (en el cas de Wikidata, el coneixement enciclopèdic que representa) creada a partir de l’acord entre diferents individus, i expressada amb un llenguatge formal. L’ontologia aporta significat als diferents elements que es combinen en les declaracions, i ho fa formalment per tal que pugui ser processat i comprès no únicament per humans sinó també per enginyers informàtics. Una ontologia té tres components fonamentals. Un ja l’hem vist en descriure l’expressió de les declaracions: les propietats. Els altres dos components són: les classes generades a partir de l’agrupació d’elements individuals (coses, persones, idees...) que comparteixen característiques, i que es relacionen entre elles en forma de taxonomia; i les restriccions semàntiques, que són regles sobre com s’apliquen les propietats i les classes per a la descripció d’elements individuals. En el següent apartat ens aproximem a les característiques més rellevants de l’ontologia de Wikidata, i als elements que la singularitzen.

6.1.2. Ontologia en el context de Wikidata

L’ontologia de Wikidata és de tipus superior (en anglès, *upper ontology*); és a dir aplicable als diferents dominis del coneixement. Des del punt de vista de la seva creació i desenvolupament, es tracta d’una ontologia “controlada per la comunitat”; això significa que no parteix d’una predefinició d’especialistes ni es basa en una de preexistent. En el moment actual, tampoc l’ontologia segueix completament els principis del web semàntic, ni està dissenyada seguint de forma estricta els seus estàndards. En els darrers temps, però, s’observa un procés d’aproximació a aquests estàndards, amb els quals ja s’han fet mapatges⁶².

⁶² Els mapatges es poden consultar en Wikidata. (2022). Wikidata:Relation between properties in RDF and in Wikidata. https://wikidata.org/wiki/Wikidata:Relation_between_properties_in_RDF_and_in_Wikidata

Els components fonamentals de l'ontologia són les **propietats**, a les que ja hem fet referència en l'apartat com a components de les declaracions, i les **classes**. Una propietat és un descriptor o atribut per al valor d'una declaració. Per exemple, la propietat *sexe o gènere* és el descriptor que permet atribuir el valor *gènere no-binari* a l'element *Bel Olid Báez*. Una classe és una agrupació. Les classes són agrupacions d'elements individuals definides a partir de les característiques comunes als elements individuals. Per exemple, l'element individual *Bel Olid Báez* forma part de la classe *humà*, i comparteix les característiques de tots els elements individuals que en formen part. En el llenguatge propi de les ontologies es diu que *Bel Olid Báez és una instància de la classe humà*.

A més, les propietats i classes estan marcades amb una sèrie de **restriccions semàntiques** que, d'antuvi, estableixen les condicions de la seva aplicació, i, a més, juguen un rol clau en processos més avançats de **proves de consistència i inferència**. Els primers asseguren que les declaracions generades sobre un element són correctes des del punt de vista semàntic; és a dir, són certes. La inferència permet la generació automàtica de noves declaracions a partir de l'aplicació de regles d'implicació (*si A, llavors B*) sobre declaracions explícites, efectivament creades i publicades en la base de dades de Wikidata.

A continuació es detallen les condicions de creació d'aquests components en l'ontologia de Wikidata, s'exemplifiquen a partir de l'àmbit del gènere, i es valoren a partir de mostres i observació. A més, identifiquem els àmbits de millora que cal abordar, tant per a l'explotació de l'ontologia en el context de Wikidata, com en una futura aplicació a Wikipedia.

6.1.2.1. Propietats

A Wikidata, poden crear propietats els usuaris a qui es reconeix la capacitat tècnica de fer-ho. També tenen aquest dret, per defecte, els administradors. D'acord amb la directriu "Wikidata:Creatores de propietats"⁶³, del 6 de març de 2023, hi havia 116 creadors de propietats, de què 65 eren administradors.

Les condicions per a la creació d'una propietat es detallen en Wikidata:Property creation/es⁶⁴. Resulta especialment significatiu que la proposta ha d'haver estat feta per algú diferent del que finalment la crea, i la proposta ha de tenir suficient consens a favor, la qual cosa significa que ha de tenir en compte la reflexió i la lògica dels punts de discussió. I, de fet, una veu en contra amb suficient reflexió pot bloquejar la creació de la propietat.

Les propietats són descrites amb el mateix tipus de declaracions que hem exposat en l'apartat 6.1.1. per als elements de Wikidata, i que hem exemplificat en el cas de *Bel Olid Báez* (Q16176248). Les propietats, però, tenen una especificitat molt important que són les declaracions de restriccions.

Una **restricció** és una regla sobre com s'ha d'utilitzar una propietat determinada. En l'actualitat, hi ha 44 tipus de restriccions, que es poden consultar en Help:Property constraints portal/list of constraints⁶⁵. Cada un d'aquests tipus és un element (classe) de Wikidata, i, com qualsevol element, és descrit mitjançant declaracions. Veiem un parell d'exemples.

⁶³ Wikidata. (2023, març 6). *Wikidata:Creatores de propietats*. https://wikidata.org/wiki/Wikidata:Property_creators/es. És una política de Wikidata.

⁶⁴ Wikidata. (2023, abril 24). *Wikidata:Property creation/es*. https://wikidata.org/wiki/Wikidata:Property_creation/es. És una política de Wikidata.

⁶⁵ Wikimedia. (2023, juny 6). *Help:Property constraints portal/list of constraints*. https://wikidata.org/wiki/Help:Property_constraints_portal/list_of_constraints

- La *restricció de tipus* (Q21503250) permet especificar quin tipus d'entitat pot ser subjecte de la propietat⁶⁶. Els tipus d'entitats corresponen a classes de l'ontologia.
- La *restricció un-de* (Q21510859) permet especificar que el valor de la propietat s'ha d'escollir entre un conjunt delimitat d'elements.

En el model de dades de Wikidata, les restriccions es declaren mitjançant tripletes formades pels següents components.

- El subjecte de la tripleta és la propietat sobre la qual actua la restricció. Per exemple, la propietat *sexe o gènere* (P21).
- La propietat de la declaració és *restricció* (P2302).
- L'objecte de la declaració és un dels 44 elements a què hem fet referència abans.

La literalitat de les declaracions de restriccions relatives a la propietat *sexe o gènere* (P21) són les següents.

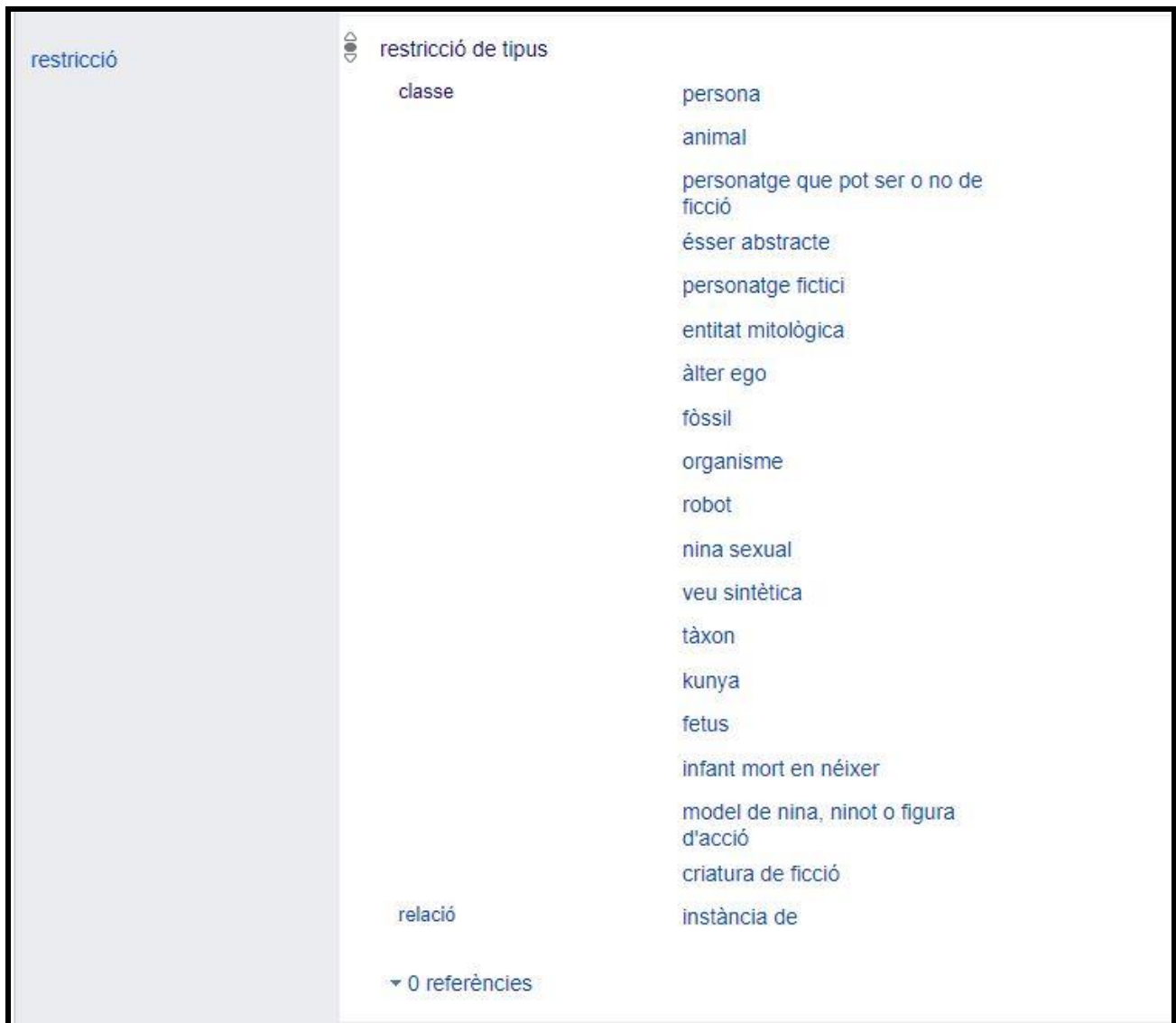
- Declaració 1: La propietat *sexe o gènere* (P21) té una *restricció* (P2302) del tipus *restricció de tipus* (Q21503250).
- Declaració 2: La propietat *sexe o gènere* (P21) té una *restricció* (P2302) del tipus *restricció un-de* (Q21510859).

Normalment és necessari realitzar declaracions sobre aquestes declaracions de restricció per a detallar-ne aspectes com el seu abast, si la restricció és de caràcter obligatori o opcional, i, com hem vist en el cas dels elements, les referències de verificació. Es tracta d'un procés de recursivitat, on la declaració de restricció es transforma, ella mateixa, en un element, de forma que pot formar ocupar la posició de subjecte d'una declaració, i ser descrit per altres declaracions.

En la Imatge 14 observem la secció de la pàgina sobre *sexe o gènere* on es presenta la Declaració 1. Aquesta declaració ha estat reificada, i és el subjecte de la propietat *classe* (P2308). I l'objecte d'aquesta propietat és el conjunt d'entitats (classes) les instàncies de la qual poden incorporar la propietat *sexe o gènere*.

⁶⁶ Correspon a la restricció de domini en el context del web semàntic, que s'expressa mitjançant la propietat *domain* del vocabulari RDFS.

Imatge 14. Restricció de tipus de la propietat sexe o gènere



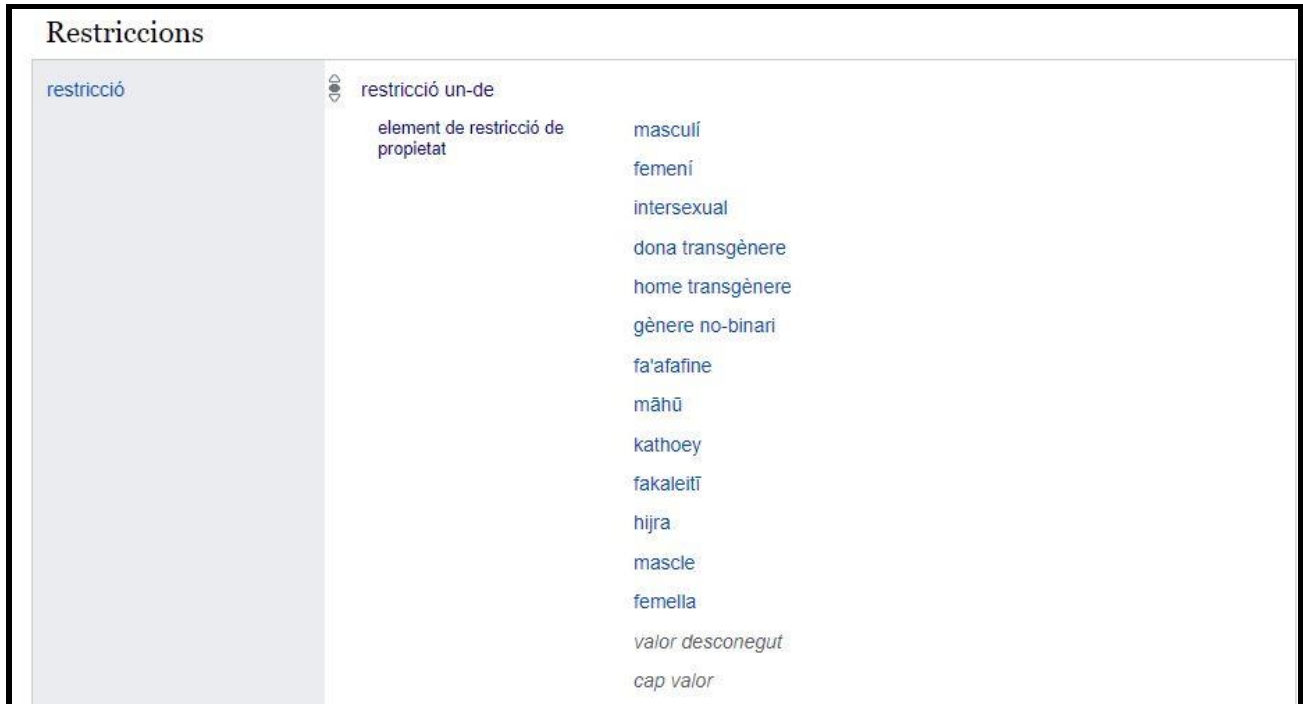
Com podem observar en la imatge, hi ha 18 classes les instàncies de les quals poden ser subjecte de la propietat *sexe o gènere*:

1. persona
2. animal
3. personatge que pot ser o no de ficció
4. ésser abstracte
5. personatge fictici
6. entitat mitològica
7. àlter ego
8. fòssil
9. organisme
10. robot
11. nina sexual
12. veu sintètica
13. tàxon
14. kunya
15. fetus

16. infant mort en néixer
17. model de nina, ninot o figura d'acció
18. criatura de ficció

Per la seva banda, en la Imatge 15 observem la secció de la pàgina sobre sexe o gènere on es presenta la Declaració 2. Aquesta declaració ha estat reificada, i és el subjecte de la propietat *element de restricció de propietat* (P2305). I l'objecte d'aquesta propietat és cada un dels elements que poden ser valor de la propietat *sexe o gènere* (P21).

Imatge 15. Restricció un-de de la propietat sexe o gènere



La llista completa està formada per 53 elements, a saber:

1. masculí
2. femení
3. intersexual
4. dona transgènere
5. home transgènere
6. gènere no-binari
7. fa'afafine
8. mähū
9. kathoey
10. fakaleitī
11. hijra
12. mascle
13. femella
14. valor desconegut
15. cap valor
16. dos esperits
17. transmasculí
18. transfemení

19. muxe
20. organisme intersexual
21. agènere
22. genderqueer
23. gènere fluid
24. neutre
25. eunuc
26. pangènere
27. *cogenitor*
28. sexe neutre
29. hermafroditisme
30. dona cisgènere
31. home cisgènere
32. tercer sexe
33. gènere X
34. *demiboy* anglès
35. *demigirl*
36. bigènere
37. persona trans
38. transvestit
39. 'akava'ine
40. assignat com a dona en néixer
41. assignat com a home en néixer
42. andrògin
43. *yinyang ren*
44. *boi*
45. persona intersexual
46. ginandromorf
47. Takatāpui
48. *undisclosed gender*
49. Fakafifine
50. fakafifine
51. gènere determinat pel jugador
52. *gender not disclosed in work*
53. *androgynos*

Wikidata manté diferents eines per a detectar les violacions de les restriccions de propietats⁶⁷, i facilitar la seva resolució. Tanmateix, i tal com hem indicat a l'inici d'aquest apartat, la correcta aplicació es deixa a la voluntat dels editors, de forma que l'explotació d'aquests elements de l'ontologia per a processos de validació de consistència i d'inferència de nou coneixement resta força limitat. I és un dels àmbits de millora urgents que han plantejat els responsables de l'ontologia, amb mesures concretes que descriurem en els apartats posteriors.

6.1.2.2. Classes i instàncies

Wikidata estableix una clara diferència entre instàncies (o individus) i classes.

- Una instància representa un concepte o objecte individual, clarament identificable.
- Una classe és un objecte abstracte que representa un conjunt d'instàncies.

⁶⁷ *Category:Constraint violation reports*. (2022, febrer 9).
https://wikidata.org/wiki/Category:Constraint_violation_reports

Tal com indica *Help:Propietats*⁶⁸ bàsiques de pertinença, normalment, totes les instàncies que pertanyen a una classe comparteixen un conjunt de propietats, propietats que caracteritzen la classe. Les instàncies difereixen entre si pels valors que tenen en aquestes propietats, però no pel fet de tenir les propietats en si mateixes. Per tant, cada classe es caracteritza normalment per les propietats que comparteixen totes les seves instàncies (tot i que Wikidata no ho aplica estrictament). L'ítem corresponent a classe de Wikidata (Q16889133) la defineix com un "conjunt d'elements que comparteixen una propietat determinada".

Una especificitat de Wikidata, però, és la possibilitat que un element sigui una instància i una classe al mateix temps. En la pràctica, però, no hi ha cap impediment que un element sigui alhora una instància i una classe. En la seva pàgina d'ajuda, Wikidata:WikiProject Ontology/Classes⁶⁹, la definició és pragmàtica.

- Un element és una instància simplement pel fet de ser el subjecte d'una propietat instància de (P31). Per exemple, l'element *gènere no-binari* (Q48270) és una instància de la classe *identitat de gènere* (Q48264). I l'element *Bel Olid Báez* (Q16176248) és una instància de la classe *gènere no-binari* (Q48270).
- Un element és una classe simplement pel fet que és l'objecte de la propietat *instància de* (P31), o el subjecte o objecte de la propietat *subclasse de* (P279). Per exemple, el mateix element *gènere no-binari* (Q48270) és una subclasse de *gènere* (Q48277), *terme genèric* (Q210588) i *minoría de gènere* (Q11894636); i superclasse de 30 classes, entre elles, *kathoey* (Q746411) o *agènere* (Q505371). En canvi, l'element *Bel Olid Báez* (Q16176248), que representa un objecte únic en el món, no és subclasse de *gènere no-binari* (Q48270) - ni de qualsevol altra classe.

La suma de les declaracions sobre ítems, i les declaracions sobre les propietats i classes que les fonamenten, constitueixen el graf de coneixement de Wikidata.

Les propietats *subclasse de* i *instància de* són les arestes fonamentals del graf de coneixement generat pels components de Wikidata. Una tercera propietat que completa la conformació del graf, de forma subsidiària, és la propietat *part de*. Aquesta propietat pot tenir com a subjecte un element únicament instància, o un element que sigui instància i classe. Per exemple, la classe *gènere no-binari* (Q48270) és part de la classe *transidentitat* (Q1771029).

L'estàndard de referència en el disseny i desenvolupament de classes es basa en RDFS (que subsumeix l'estàndard bàsic del model RDF), tot i que de forma lleugera.

Les propietats *instància de* (P31) i *subclasse de* (P279) entronquen, respectivament, amb les propietats *rdf:type* i *rdfs:subClassOf* amb un significat homologable.

- *rdf:type* = El subjecte és una instància d'una classe.
- *rdfs:subClassOf* = El subjecte és una subclasse d'una classe.

Com succeeix en els estàndards del web semàntic, totes les instàncies d'una classe donada també són instàncies de les superclasses d'aquesta classe. És a dir, com que *Bel Olid Báez* (Q16176248) és una instància de la classe *gènere no-binari* (Q48270), també és una instància de les superclasses *gènere* (Q48277), *minoría de gènere* (Q11894636), i *terme genèric* (Q210588). I, ascedint en les cadenes jeràrquiques d'aquestes tres classes, arribarem a la declaració final que

⁶⁸ Wikimedia. (2023, maig 1). *Help:Propietats bàsiques de pertinença*. https://wikidata.org/wiki/Help:Basic_membership_properties/ca. És una ajuda.

⁶⁹ Wikidata. (2022). Wikidata:WikiProject Ontology/Classes. https://wikidata.org/wiki/Wikidata:WikiProject_Ontology/Classes

Bel Olid Báez (Q16176248) és una instància de la classe arrel de l'ontologia de Wikidata, *ens* (Q35120), descrita com “alguna cosa que existeix”.

L'adscripció als estàndards del web semàntic és, però, “lleugera”; tal com s'indica en Wikidata:Item classification⁷⁰. I una mostra d'aquesta lleugeresa és, precisament, la promiscuïtat entre instàncies i classes a què hem fet referència.

Si bé l'ontologia de Wikidata no adhereix a l'estàndard OWL, en la creació i desenvolupament de les classes podem observar l'aplicació d'algunes restriccions semàntiques equiparables a aquell llenguatge. Per exemple, la propietat *subclasse de* P279 és instància de la classe *propietat transitiva* (Q18647515). A partir d'aquesta definició, si A és una subclasse de B, i B és una subclasse de C, llavors s'infereix que A és subclasse de C. Aquest axioma permet generar jerarquies de classes des del node superior (entitats més generals) fins els nodes inferiors (entitats més específiques), molt útils en contextos de navegació i cerca per exploració o per interrogació.

Aquesta inferència es pot generar en entorns on les declaracions estiguin acompanyades de regles d'inferència efectives.

Wikidata incorpora alguns instruments en aquesta línia. En concret:

- Classe *entitat fonamental de Wikidata* (Q115490628), que en el moment actual incorpora 11 instàncies directes, entre elles la classe *propietat transitiva* (wd:Q18647515).
- Classe *propietat de Wikidata per la relació entre classes* (Q28326461), que en el moment actual incorpora 5 instàncies directes, entre elles la *propietat subclasse de* (P279).

En conjunt, però, el nombre de definicions de restriccions sobre classes per a la generació d'inferències i control de consistència amb recursos propis és limitat; incomparablement inferior al que ofereixen els estàndards del web semàntic.

A diferència del que passa amb les propietats, la creació de classes pot ser assumida per qualsevol persona editora. Possiblement, aquesta apertura proporciona un enriquiment de les perspectives amb què es crea l'ontologia de Wikidata; però també té els seus riscos, com veurem en l'avaluació a partir d'indicadors que presentem a continuació.

Els responsables del desenvolupament de l'ontologia de Wikidata i, de fet, del graf de coneixement en el seu conjunt, són conscients de la necessitat d'introduir millores en el seu procés de creació. S'han creat diferents recursos per a facilitar l'ús de l'ontologia en l'edició i enriquiment d'entitats. Alguns d'aquests recursos estan orientats a la consulta en el moment d'editar o enriquir entitats. En són exemples les ajudes com *Help:Property constraints portal*⁷¹, on s'exposen les regles sobre les propietats que determinen com s'han d'utilitzar; i les llistes de propietats recomanades per experts per a descriure tipus d'entitats específics (per exemple, *Wikidata:WikiProject Biography*⁷² per a la descripció de biografies).

Hi ha, però, un recurs que guia la tasca d'edició de categories d'entitats específiques, i prescriu els tipus, propietats i valors a assignar. Es tracta dels *Entity Schemas*⁷³, formalment expressats

⁷⁰ Wikidata. (2020, gener 21). *Wikidata:Item classification*. https://wikidata.org/wiki/Wikidata:Item_classification. És un treball en curs.

⁷¹ Wikimedia. (2023, març 23). *Help:Property constraints portal*. https://wikidata.org/wiki/Help:Property_constraints_portal

⁷² Wikimedia. (2022, novembre 30). *Wikidata:WikiProject Biography*. https://wikidata.org/wiki/Wikidata:WikiProject_Biography

⁷³ Wikimedia. (2023, juny 25). *Wikidata:WikiProject Schemas*. https://wikidata.org/wiki/Wikidata:WikiProject_Schemas#Tools

amb un llenguatge específic (Shape Expression language o ShEx). Se n'han creat per a un nombre creixent de tipus d'entitats; entre elles, les persones. Per a la creació d'aquesta categoria d'entitats, disposem, actualment, de dos esquemes.

- Entity Schema E10⁷⁴, per a descriure instàncies de la classe humà (Q5) mitjançant 22 propietats.
- Entity Schema E14 (*basic properties*)⁷⁵, per a descriure instàncies de la classe humà de forma més simple, mitjançant únicament sis propietats.

És important assenyalar que ambdós esquemes incorporen la propietat *sexe o gènere* (P21) de forma prescriptiva.

Per a l'aplicació dels esquemes, els editors disposen d'eines de generació d'elements a partir de dades en format csv, validadors d'entitats generats a partir d'esquemes, etc. En conjunt, aquest projecte vol millorar la qualitat de les dades publicades, i la usabilitat d'aquestes dades en els contextos de cerca, a més d'impulsar el creixement de la comunitat d'editors mitjançant la simplificació dels processos i la reducció dels conflictes.

6.1.2.3. Etiquetatge d'elements i propietats

Els elements i propietats de Wikidata disposen d'expressions en llenguatge natural que, en conjunt, són designades termes. Existeixen tres tipus de termes: les etiquetes, o designacions principals per a identificar les entitats; els àlies, o designacions secundàries; i les descripcions, o textos explicatius sobre les entitats. L'objectiu de Wikidata és que aquests components estiguin disponibles en totes les llengües.

Les etiquetes són especialment rellevants en els contextos de cerca. Únicament pot haver-hi una per llengua. És l'instrument de comunicació entre la necessitat de la persona usuària i les entitats que conformen l'ontologia. De la major o menor precisió i exhaustivitat amb què s'hagin designat les entitats dependrà l'èxit dels processos de cerca dels usuaris en les seves diverses modalitats (navegació, exploració i interrogació).

L'usuari Mr. Ibrahim publica les estadístiques de percentatge d'etiquetes d'elements disponibles per llengua⁷⁶. El 6 de març de 2023 (darrer registre en el moment de redactar aquest informe), el percentatge d'etiquetes en llengua catalana era del 14,9% (en xifres absolutes, 15.116.853 de 101.045.411 possibles). En aquesta data, ocupava la vuitena posició de 539 llengües. Val a dir que si bé les dues primeres llengües, l'anglès i el neerlandès, assoleixen graus de disponibilitat d'elements destacables, el 85,8% i el 62,3%, respectivament, a partir de la tercera posició, el percentatge de disponibilitat cau fins al 24,2%.

En canvi, la situació de les propietats és molt més avantatjosa per a la llengua catalana. El 29 de juny de 2023 n'hi havia 11.051 a Wikidata; i d'aquestes, el 98,20% estaven disponibles en català (en xifres absolutes, 10.853).

Una altra dimensió de les etiquetes rellevant en el marc d'aquest informe és el seu grau d'alineament pel que fa a la perspectiva de gènere. Els creadors i editors de propietats disposen d'una pàgina d'ajuda, *Help:Label/ca*⁷⁷, amb l'advertiment que és una traducció de les propostes per a l'anglès, i únicament el “*punt de partida per a futures directrius pròpies de les etiquetes en*

⁷⁴ <https://wikidata.org/wiki/EntitySchema:E10>

⁷⁵ <https://wikidata.org/wiki/EntitySchema:E14>

⁷⁶ *User:Mr. Ibrahim/Language statistics for items*. (2023, juny 18).

https://wikidata.org/wiki/User:Mr. Ibrahim/Language_statistics_for_items.

⁷⁷ Wikimedia. (2016, desembre 29). *Help:Label/ca*. <https://wikidata.org/wiki/Help:Label/ca>.

català”. La versió original en anglès⁷⁸ sí incorpora una directriu alineada amb la perspectiva de gènere.

“If the label is different in the male form and in the female form, it is recommended to use a gender neutral form if it exists in common usage (i.e. use a form that applies to males and females). Avoid using the male form as a generic form when possible. Examples: bomber (Q107711) : firefighter is gender neutral and **fireman** is a male form. It is better to use **firefighter**.”

En canvi, en la versió en català, aquesta directriu no ha estat incorporada en la traducció. De fet, les etiquetes formades pels noms derivats amb el sufix -tor/-tora adopten únicament la forma en masculí: autor, compositor, director, editor, elector, gerent/director, locutor, productor, rector, traductor... i en tots els termes compostos amb aquestes paraules.

Aquests trets de les etiquetes i, específicament, de les que es generen per a la designació de les propietats tenen conseqüències molt important en els serveis de cerca de Wikidata, com veurem en els apartats que segueixen.

6.1.3. El fons de dades de Wikidata

Les dades de Wikidata poden ser utilitzades per diferents propòsits. Més enllà de la consulta puntual d'un element o d'un conjunt d'elements, Wikidata posa a disposició dels usuaris mètodes d'accés per a l'enllaçament de dades sense haver de descarregar-les en un altre servidor, per a l'enriquiment de dades de tercers, o per a la generació de serveis de cerca locals. En tots els casos, el consum de les dades de Wikidata pot ser realitzada per usuaris humans, o per sistemes automàtics (bots)⁷⁹.

En una de les guies de Wikidata, Accés a les dades⁸⁰, s'identifiquen i descriuen vuit mètodes d'accés a les dades de Wikidata, quatre dels quals s'orienten a la interacció directa amb usuaris, que necessiten recuperar quantitats limitades de resultats.

- Cerca⁸¹ per a contextos on necessitem cercar una cadena de text, quan coneixem possibles designacions de les entitats, o quan podem especificar la nostra consulta a partir de relacions molt simples entre les dades.
- Linked Data Interface (URI)⁸² per a contextos on necessitem recuperar entitats individuals i completes que ja coneixem.
- Servei de Consulta de Wikidata⁸³ per a contextos on coneixem el patró bàsic d'estructuració de dades en tres components: l'entitat que és subjecte del nostre interès en primera posició (subjecte); la propietat que descriu l'entitat en posició intermèdia; i el valor de la propietat per a l'entitat en tercera posició (objecte). Proporciona dues interfícies: una que exigeix el coneixement del llenguatge i protocol de consulta **SPARQL**,

⁷⁸ Wikidata. (2023). *Help:Label*. <https://wikidata.org/wiki/Help:Label>. En el moment d'elaborar aquest informe era una proposta de norma o directriu.

⁷⁹ <https://wikidata.org/wiki/Wikidata:Bots>

⁸⁰ Wikimedia. (2023, març 28). *Wikidata:Accés a les dades*. https://wikidata.org/wiki/Wikidata:Data_access/ca

⁸¹ Wikimedia. (2023). *Cerca*. <https://wikidata.org/wiki/Special:Search>

⁸² Wikimedia. (2022, setembre 26). *Vikidata:SPARQL Servei de consulta/Vikidata Ajuda de Consulta*. https://wikidata.org/wiki/Wikidata:SPARQL_query_service/Wikidata_Query_Help/ca.

⁸³ <https://query.wikidata.org/>

i una altra que facilita la generació de consultes en aquest llenguatge mitjançant una interfície assistida.

Tots els mètodes d'accés a les dades de Wikidata operen a partir d'un fons format pel sistema de gestió de dades RDF, o repositori RDF, Blazegraph. El 23 de febrer de 2022, aquest sistema ha de gestionar les següents magnituds, d'acord amb Vrandečić et al. (2023).

- Col·laboradors:
 - registrades 565.000
 - no registrats (IP diferents) 1,6 milions
 - actius al mes > 46.000
 - bots: 359⁸⁴
- Nombre d'elements >101 milions
- Nombre de propietats >10.800
 - per a identificadors externs > 7.800
- Nombre de declaracions >1.440 milions
 - per a identificadors externs >206 milions
 - mitjana per elements 14,3
- Nombre d'edicions > 1.800 milions
 - per dia 699.000
- Visualitzacions mensuals de pàgines (mitjana de 12 mesos): 420 milions
- Articles de Wikipedia que utilitzen Wikidata: 75-97% (gener 2023)
- Articles de Viquipèdia que utilitzen Wikidata (caWiki): 90,6% (gener 2023, inclou les infotaules dels articles, auto-categorització, descripcions i senyalització de treballs de manteniment)

Indubtablement, aquestes xifres són impressionants, i són les pròpies de la base de dades secundària i oberta més gran existent en l'actualitat. Això no obstant, en els darrers anys les valoracions sobre el grau de compliment de processos, i sobre accessibilitat i ús dels serveis de cerca, ha mostrat signes preocupants d'estancament.

Els responsables de Wikidata són plenament conscients d'aquestes limitacions i, de fet, han posat la mirada en la necessitat de substituir el programari de fons de Wikidata, Blazegraph⁸⁵, per un altre que pugui respondre millor als reptes de creixement i de qualitat. Han identificat dos problemes principals pel que fa a l'eficiència d'aquest programari.

- Blazegraph és una implementació d'un sol servidor (amb rèplica per a la disponibilitat) que pateix problemes de rendiment amb la mida actual de la base de dades d'aproximadament 14 bilions de tripletes o declaracions.
- El rendiment de les consultes es basa en una resposta síncrona i una tecnologia "antiga", i està experimentant temps d'espera i inestabilitat general.

Un argument conjuntural que ha aconsellat aquesta substitució ha estat l'adquisició Blazegraph per Amazon, i la seva intenció de cancel·lar-ne el suport i manteniment a curt termini.

L'equip de cerca de Wikidata va alertar sobre el risc de Blazegraph failure en el August 2021 scaling update⁸⁶, i, des de llavors, ha impulsat un procés obert d'identificació de problemes a resoldre i de millores a assolir, que va més enllà d'una simple substitució de programari. L'estat

⁸⁴ <https://m.wikidata.org/wiki/Wikidata:Bots>

⁸⁵ <https://blazegraph.com/>

⁸⁶ https://wikidata.org/wiki/Wikidata:SPARQL_query_service/WDQS_backend_update/August_2021_scaling_update

de la qüestió sobre aquest procés es pot consultar en el document de treball WDQS Search team (2022)⁸⁷. El compromís és donar resposta a aquests reptes com més tard el 2026.

Amb relació a les inconsistències en l'ontologia a què hem fet referència anteriorment, els requeriments d'avaluació establerts incorporen l'aplicació de llenguatges orientats a la integrity checking més avançats que les funcions d'SPARQL. En concret, l'informe WDQS fa referència a **Shapes Constraint Language** (a partir d'ara, SHACL). SHACL permet la validació de gràfics i incorpora no només la capacitat d'especificar un nivell de gravetat dels resultats de la validació, sinó també la possibilitat de retornar suggeriments sobre com es poden arreglar les dades si es produeix el resultat de la validació.

6.2. Avaluació del sistema d'ontologies per a la navegació i cerca

En aquest apartat s'aporten els resultats de l'avaluació del sistema de navegació i la cerca de continguts a Viquipèdia a través del sistema d'organització del coneixement de Wikidata (ontologies). L'avaluació s'ha fet segons dos mètodes d'avaluació, d'una banda l'avaluació amb persones usuàries (lectores i editores) en el qual el focus de l'avaluació és en la usabilitat i l'experiència d'usuari, i de l'altra, l'avaluació amb experts a través de l'inspecció d'estàndards i avaluació heurística.

6.2.1. Avaluació amb persones usuàries

A l'apartat de metodologia (2.3.2.) s'especifica les tasques i la mostra (taules 6 i 7) dels experiments que es realitzen amb persones usuàries. En concret per avaluar el sistema

La tasca 9 sobre el generador de consultes amb QueryBuilder es va poder realitzar a tota la mostra de 20 persones, però la tasca 10, de cercar entre el llistat de propietats de Wikidata, es va realitzar només amb 5 persones editores per la sofisticació de la tasca.

6.2.1.1. Eficiència

L'experiment per avaluar l'eficiència del sistema de navegació i cerca es va realitzar en la tasca 9 que permetia mesurar l'eficiència del cercador de Wikidata QueryBuilder. La tasca es va encarregar a les 20 persones de la mostra. En canvi, per a les tasques de navegació amb les llistes de propietats de Wikidata (tasca 10), que suposen un coneixement més especialitzat, només es van fer a 5 dels 20 usuaris, les persones editores que durant el test van demostrar expertesa en el camp dels llenguatges controlats.

En la tasca 9 en les consultes de Wikidata, es va observar una mitjana de 3,7 intents per usuari, la qual indica que cada persona usuària va haver de realitzar diverses proves o ajustos per obtenir els resultats desitjats (Vegeu taula 20), malgrat que la tasca es podia resoldre amb un sol intent (recorregut cognitiu). Aquesta dada indica que l'índex d'eficiència és del 27%. Cal destacar que la moda estadística és d'1 sol intent.

Taula 20. Índex d'eficiència del cercador de Wikidata 'QueryBuilder'

Mitjana d'intents dels usuaris	Intents mínims necessaris	Eficiència de navegació
3,7	1	27%

⁸⁷ WDQS Search team. (2022). *WDQS Backend Alternatives: The process, details and results* [Working paper about the Wikidata Query Service backend alternatives].

En la tasca 10 per la navegació de les llistes de propietats de Wikidata, la mitjana de clics per aquests usuaris experts (5 editors dels 10 totals testejats) va ser de 9,2. Sent 3 clics la mitjana de clics necessaris, que dividits per 9,2 (mitjana de clics realitzats), és del 22% d'eficiència (Vegeu taula 21). Pel que fa a la probabilitat d'error, la mitjana del retorn va ser de 3,8 i la mitjana de clics realitzats va ser de 9,2. Per tant, la probabilitat d'error és del 41%.

Taula 21. Índex d'eficiència i probabilitat d'error a la navegació per les llistes de propietats de Wikidata

Mitjana de clics fets	clics mínims necessaris	eficiència de navegació	Mitjana de retorns	Probabilitat d'error
9,2	3	33%	3,8	41%

Tenint en compte aquests índexs, podem concloure que si bé l'índex d'eficiència és bo en el cercador i les llistes de Wikidata, la probabilitat d'error encara és alta i pot suggerir que hi ha marge de millora per facilitar la navegació i reduir la probabilitat d'error. Es podria optimitzar la interfície seguint un procés d'UX o proporcionar guies i ajuda contextual als usuaris durant la seva interacció amb les llistes de propietats.

6.2.1.2. Efectivitat

Basant-nos en les valoracions proporcionades, es pot concloure que les opinions sobre la consulta a Wikidata són variades i sovint contradictòries. Algunes persones consideren que és complexa, caòtica i ineficient, mentre que altres la veuen com una eina rellevant, senzilla i amb un llenguatge controlat. També hi ha mencions a l'ús d'eines externes per complementar la consulta i a la necessitat de millorar-la perquè sigui més precisa i intuïtiva, especialment per a aquells que no són experts en l'àmbit.

6.2.1.3. Facilitat d'aprenentatge

Les persones usuàries han expressat una valoració relativament baixa d'1,35 sobre 5 en termes d'intuïtivitat del cercador de Wikidata. Aquesta puntuació es correspon amb els comentaris anteriors, on s'ha assenyalat que el cercador pot ser considerat una mica caòtic, amb explicacions poc específiques i una selecció de paraules per a les cerques que pot ser rebuscada.

També han mencionat que l'intuïtivitat del cercador pot variar en funció del coneixement prèviament adquirit sobre les propietats de Wikidata i que l'ús de llenguatge natural està limitat. Algunes opinions indiquen que, tot i ser més intuïtiu que el cercador de Viquipèdia, encara hi ha marge per a la millora.

En resum, l'intuïtivitat del cercador de Wikidata ha estat percebuda com a insatisfactòria pels usuaris amb una puntuació baixa i els comentaris anteriors reafirmen aquesta opinió. És important que Wikidata tingui en compte aquests comentaris per tal de millorar l'experiència d'ús i proporcionar una interfície més clara i intuïtiva per a les seves persones usuàries.

6.2.1.4. Trobabilitat

Com els anteriors indicadors, diferenciem les consultes a Wikidata (tasca 9) i de la navegació de les llistes de propietats (tasca 10).

Pel que fa als editors, el 90% d'èxit indica que un 90% dels editors que van utilitzar les consultes de Wikidata van aconseguir els resultats esperats. Això suggereix que la majoria dels editors van poder obtenir la informació necessària per a les seves tasques d'edició amb èxit (Vegeu Taula 22).

Però, en el cas dels lectors, només un 40% dels usuaris van aconseguir els resultats desitjats a través de les consultes de Wikidata. Aquesta baixa taxa d'èxit pot ser atribuïda a diversos factors, com ara la manca d'experiència dels usuaris o la complexitat de les consultes realitzades (Vegeu Taula 22).

Taula 22. Nombre de persones que satisfan la necessitat informativa amb el cercador de Wikidata

Persones usuàries en global	65%
Grup de persones editores	90%
Grup de persones lectors	40%

En general, aquests indicadors d'èxit ens mostren que hi ha una certa eficàcia en l'ús de les consultes de Wikidata per part dels editors, ja que la majoria aconseguen els resultats desitjats. No obstant això, els lectors en general tenen una taxa d'èxit més baixa, el que indica la necessitat de millorar l'experiència d'usuari i proporcionar més suport o recursos per ajudar-los a aconseguir els resultats esperats a través de les consultes de Wikidata.

En canvi, l'èxit de la navegació per les llistes de propietats (tasca 10) és d'un 80%, ja que 4 de 5 testejat van resoldre la tasca. Amb tot, això no pot negar la dificultat de la interfície per trobar les propietats i la necessitat d'un coneixement previ per a entendre la col·locació de les propietats i els seus valors per a saber on es localitzen correctament els continguts.

Aquests resultats apunten a un marge de millora important si s'empren mètodes per redissenar les opcions de cerca i navegació de Wikidata amb una perspectiva de disseny centrat en les persones (UCD)⁸⁸.

6.2.1.5. Satisfacció

Les persones usuàries han valorat la seva satisfacció del cercador de QueryBuilder a Wikidata amb una puntuació de 0 a 5, amb el resultat d'una mitjana de l'1,55. Aquest valor és més baix que els cercadors de la Viquipèdia, encara que els comentaris fets pels testejats són diverses i qualifiquen Wikidata com una eina amb molt més potencial que en la cerca de Viquipèdia. Algunes d'elles destaquen la percepció de caos i la falta de claredat en les explicacions del funcionament de Wikidata. S'assenyala que, tot i ser més intuïtiu que el cercador de Viquipèdia, encara pot ser complicat realitzar cerques i seleccionar les paraules adequades.

A més, s'indica que l'ús de llenguatge natural és limitat i que cal tenir coneixement prèviament sobre les propietats de Wikidata per utilitzar-lo adequadament. Alguns usuaris també mencionen la ubicació del cercador, que no és prou visible.

⁸⁸ User Centered Design. <https://www.interaction-design.org/literature/topics/user-centered-design>

En general, es reconeix que Wikidata té potencial per a millorar i esmenten la importància de conèixer les propietats abans d'utilitzar la plataforma per obtenir resultats satisfactoris. Aquests comentaris reflecteixen les experiències i opinions diverses dels usuaris en relació amb l'ús i l'aprenentatge de Wikidata.

6.2.2. Inspecció d'estàndards i avaluació heurística

Els membres del projecte Ontology⁸⁹ han identificat les limitacions que presenta per a ser considerada una ontologia correcta. Aquestes limitacions es divideixen en dos grups.

- Un primer grup van ser inicialment identificats en WikidataCon 2021⁹⁰, i s'orienten a superar les barreres a la reutilització de les dades per altres serveis i projectes.
- Un segon grup són considerats com a problemes existents en la representació del coneixement per Wikidata.

En el marc d'aquest informe, ens interessa principalment el primer grup, ja que identifica elements a superar si es vol aplicar en la categorització dels continguts de Viquipèdia. A continuació, i partint de les barreres a la reutilització formulades pels membres del projecte, presentem exemples relacionats amb les classes que conformen la restricció de rang de la propietat P21 (gènere o sexe).

Els indicadors han estat seleccionats tenint en compte el seu interès per a l'informe i l'adequació a les necessitats de l'avaluador.

6.2.2.1. Instàncies utilitzades com a classes

Definició: Les declaracions amb la propietat *instància de* (P31) només poden tenir classes com a valors. Aquesta restricció es manifesta en la declaració que aquesta propietat és del tipus *Wikidata property for the relationship of the element to its class* (Q28326730).

Evidències:

- *Asunción Estévez* (Q115264435) és subclasse de *dona* (Q467); *Ilya Varlamov anglès* (Q4103885) és subclasse de *llei sobre l'agent estranger de Rússia* (Q17071473).
- Situació específica en les classes que constitueixen la restricció de rang de la propietat P31. Tendència a considerar bessons de persones de cada gènere com a subclasses d'un gènere corresponent. Estrictament parlant, no constitueixen elements-classe, i, en conseqüència, poden ser subjecte de propietats instància de o part de, però no subclasse de.

Conseqüències: Es generaran jerarquies de classes improductives des del punt de vista de la navegació i la cerca, ja que en els nodes ocupats per instàncies, la persona usuària no podrà escollir entre més d'un individu d'un mateix tipus.

6.2.2.2. Desgavell en els nivells superiors de l'ontologia

Definició: El nivell superior de l'ontologia ha d'estar format per classes que representen conceptes molt generals (per exemple, Temps, Espai, Esdeveniment...) independents de

⁸⁹ Wikimedia. (2023). *Wikidata:WikiProject Ontology*. <http://www.inf.ufes.br/~gguizzardi/wiki-www-2016.pdf>

⁹⁰ https://wikidata.org/wiki/Wikidata:WikidataCon_2021

tot domini o problema específic. Aquests conceptes fonamentals han de ser excloents entre si, i, en el seu conjunt, han de saturar els àmbits del coneixement que l'ontologia expressa.

Evidències: En el nivell superior de l'ontologia, es troben les *classes-arrel*, i, entre aquestes, la classe superior *ens* (Q35120). El primer grup incorpora 201 suposades classes, entre les quals trobem classes que no han estat vinculades a cap superclasse - per exemple, *còmic* (Q58209506)-, i instàncies de classes com *Civil Code of the Republic of Korea* (Q5124449)-. En un nivell inferior trobem la classe que sí assumeix aquest rol, *ens* (Q35120). El nombre de subclasses directes, però, és excessiu (59 el 22/6/2023), i, a més, des d'una perspectiva conceptual, no constitueixen una representació dels àmbits fonamentals del coneixement, disjunts entre ells, i saturadors en la seva suma.

Conseqüències:

- En els contextos de cerca, on calgui partir des del nivell superior de l'ontologia, o remuntar la jerarquia fins al nivell superior, constitueixen un escenari poc clar i intuïtiu per a prendre una decisió ràpida i fonamentada.
- En contextos de raonament a partir d'axiomes vinculats als components de l'ontologia, la generació d'inferències i valoracions de consistència es veuran àmpliament afectades pel desgavell en els nivells superiors.
- En contextos de cooperació i enllaçament de dades amb altres projectes, la inexistència d'un nivell superior de l'ontologia no estandarditzat ni homologable dificulta els processos automatitzats d'interoperabilitat.

6.2.2.3. Desviació semàntica

Definició: Una entitat és contemplada des de dues o més perspectives, les propietats són diferenciables en cada perspectiva, i en canvi es fusionen en una única classe. Les relacions de subclasses observades de forma individual són correctes; però la conjunció de totes les subclasse, no.

Evidències:

- La classe *minoria de gènere* (Q11894636) és observada des de dues perspectives: és subclasse indirecta d'*entitat col·lectiva* (Q99527517), i és subclasse indirecta de classes situades en l'àmbit intangible de la representació, com *objecte abstracte* (Q7184903).
- La classe *māhū*⁹¹ és observada des de dues perspectives: és subclasse indirecta d'*entitat col·lectiva* (Q99527517), i és subclasse indirecta d'*objecte abstracte* (Q7184903).

Conseqüències:

- En el context de la cerca, l'ascens per les cadenes superiors d'una classe portaran a nodes amb una distància semàntica tan gran que desbordarà la intuïció i comprensió dels usuaris.
- Aquesta mateixa conseqüència es pot produir en contextos d'inferència a partir de la transitivitat en la *subclasse de* (P279).

6.2.2.4. Cicles o bucles de la propietat subclasse de (P279)

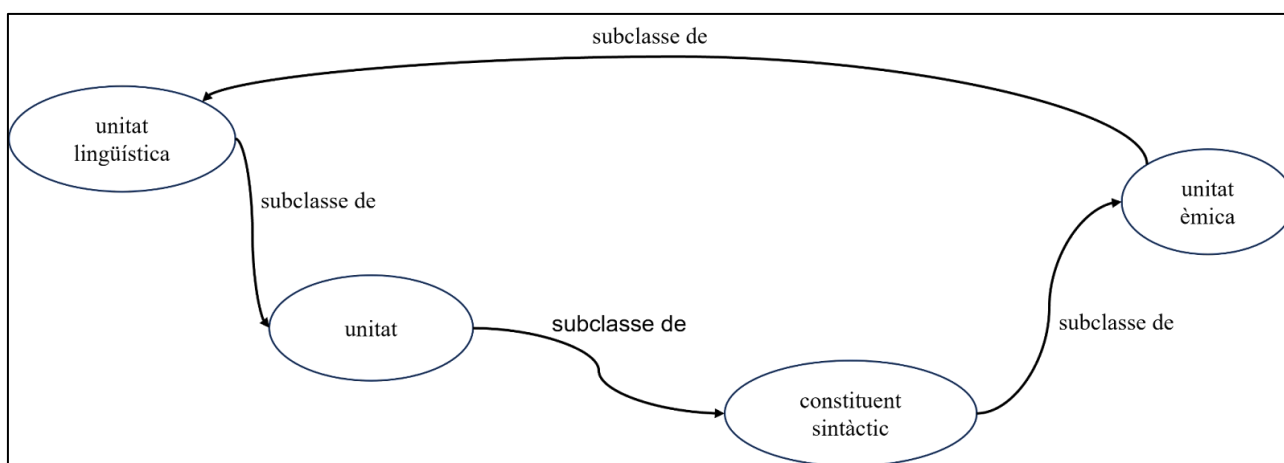
Definició: Una classe A té una subclasse directa o indirecta B, i, al mateix temps, la classe B és subclasse directa o indirecta d'A.

⁹¹ A Wikidata, aquesta classe és descrita com "tercer gènere a la cultura tradicional hawaiana i maohi".

Evidències:

- La classe *unitat lingüística* (wd:Q11953984) és subclasse indirecta de *unitat èmica* (wd:Q5371079)⁹²; i *unitat èmica* (wd:Q5371079) és subclasse directa de *unitat lingüística* (wd:Q11953984). Es representa gràficament en la Figura 5.

Figura 5. Exemple de cicle o bucle de la propietat subclasse de P279



- La classe *constituent sintàctic* (wd:Q1786828) és subclasse indirecta de *unitat* (wd:Q2198779)⁹³; i *unitat* (wd:Q2198779) és subclasse directa de *constituent sintàctic* (wd:Q1786828).

Conseqüències:

- Tant en els contextos de cerca, com en els contextos de raonament a partir d'axiomes, resultarà impossible determinar quins elements són més generals i quins elements són més específics. El bucle es traslladarà al procés de cerca.

6.2.2.5. Generalització redundant

Definició: Una classe A és subclasse de B, i, a la vegada, una subclasse d'una subclasse directa o indirecta de B.

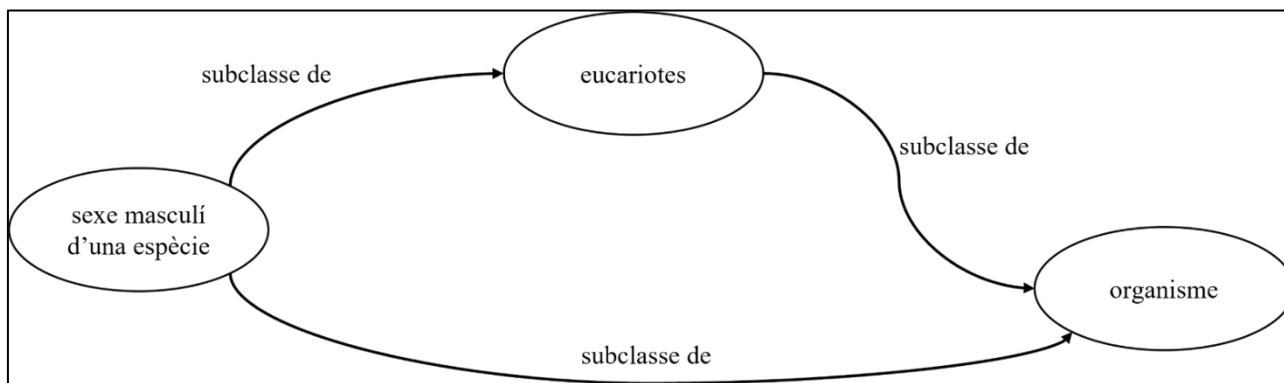
⁹² La cadena jeràrquica completa és: *unitat lingüística* (wd:Q11953984) és subclasse de *unitat* (wd:Q2198779), *unitat* és subclasse de *constituent sintàctic* (wd:Q1786828), i, finalment, *constituent sintàctic* és subclasse de *unitat èmica* (wd:Q5371079).

⁹³ La cadena jeràrquica completa és: *constituent sintàctic* (wd:Q1786828) és subclasse de *unitat èmica* (wd:Q5371079), *unitat èmica* és subclasse de *unitat lingüística* (wd:Q11953984), i *unitat lingüística* és subclasse de *unitat* (wd:Q2198779).

Evidències:

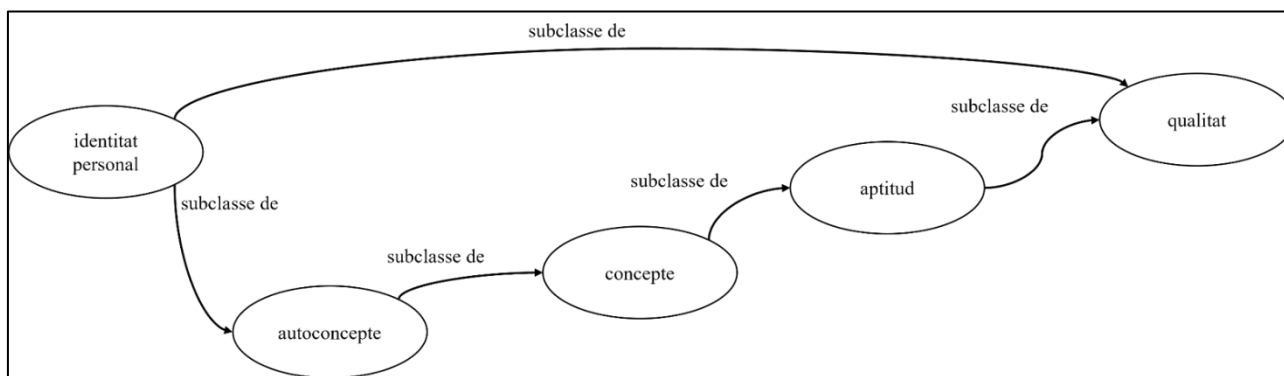
- La classe mascle (wd:Q44148) és subclasse de organisme (wd:Q7239). Alhora, mascle (wd:Q44148) és subclasse de eucariotes (wd:Q19088), i eucariotes és subclasse d'organisme (wd:Q7239). Es representa gràficament en la Figura 6.

Figura 6. Exemple de generalització redundant (I)



- La classe identitat personal (wd:Q844569) és subclasse de qualitat (wd:Q1207505). Alhora, identitat personal (wd:Q844569) és subclasse de autoconcepte (wd:Q1860557), autoconcepte és subclasse de concepte (wd:Q151885), concepte és subclasse de aptitud (wd:Q1347367), i aptitud és subclasse de qualitat (wd:Q1207505). Es representa gràficament en la Figura 7.

Figura 7. Exemple de generalització redundant (i II)



Conseqüències:

- Provocarà una redundància que complicarà els processos de cerca i d'inferència. En el cas d'optar per la ruta curta, restaran ocultes les classes intermèdies, potser decisives en el procés de cerca o inferència.

6.2.2.6. Modelatge inconsistent

Definició: Tractament diferent de dues classes pel que fa a nombre i tipus de classes amb què es relacionen.

Evidència: La classe intersexual (Q1097630) és una classe arrel; en canvi, les classes femení (Q6581072) i masculí (Q6581097) es situen en cadenes jeràrquiques de fins a 12 nivells de superclassificació.

Conseqüències: En el context de cerca, incrementa la corba d'aprenentatge de la persona usuària sobre l'estructura d'una nova classe, ja que no pot comparar-la amb una de ja coneguda amb la qual manté connexions semàntiques.

6.2.2.7. Repetició de classes

Definició: Una mateixa classe és definida dos o més cops.

Evidències:

- Dues classes etiquetades com objectes físics, tenen dos identificadors diferents: Q61961344 i Q98119401.
- Dues classes etiquetades com objecte, tenen dos identificadors diferents: Q4406616 i Q488383.
- Dues classes etiquetades com cultura tradicional, tenen dos identificadors diferents: Q4384751 i Q4461766.

Conseqüències:

- La redundància provoca desorientació dels usuaris en els contextos de cerca, i dificulta els processos de validació de declaracions, i d'inferència de nou coneixement.

6.3. Anàlisi del procés de presa de decisions

Les bases de coneixement sovint contenen informació biogràfica sobre persones, incloses dades sensibles com ara el gènere, ètnia o orientació sexual. Malauradament, els biaixos implícits en aquestes dades poden generar injustícies (Veale & Binns, 2017; Mehrabi et al., 2021) i donen lloc a aplicacions discriminatòries que sovint afecten comunitats ja marginalitzades (Buolamwini & Gebru, 2018; Bender et al., 2021). Això és especialment cert per a les comunitats que són discriminades per raó d'identitat de gènere. Els projectes digitals han lluitat per fer front a l'acceptació més àmplia de la societat pel no-binarisme (Kessler & McKenna, 1985) tenint en compte que habitualment s'han perpetuat, fins i tot amplificat esborrat col·lectius minoritzats per raons de gènere al llarg de la història (Keyes 2018).

El projecte de recerca Wikidata Gender Diversity (WiGeDi), té com a objectiu investigar el tema de la diversitat de gènere a Wikidata i un dels aspectes que han analitzat és com la comunitat de Wikidata ha gestionat l'evolució cap a una representació més diversa de les identitats de gènere, sobretot com s'ha convertit en inclusiva amb la identitat no binària, i especialment s'han centrat en analitzar les discussions de les persones usuàries sobre el tema. A partir de l'anàlisi d'un corpus de discussions a partir de tècniques de lingüística computacional han generat una narrativa al voltant dels temes relacionats amb el gènere durant els darrers nous anys en 613 discussions en anglès a Wikidata des del novembre de 2012 al setembre de 2021.

En concret l'estudi va detectar un fort ús de termes binaris en les discussions sobre famílies, en l'etiquetatge i en els noms personals. Tanmateix, l'ús del terme no-binari "sibling" per expressar vincles familiars és encoratjador, com ho són l'ús d'altres identitats com "transgender" o

“intersex”. I de forma més global, aquesta anàlisi ha descobert que les discussions sobre temes relacionats amb el gènere han canviat dràsticament durant els primers nou anys del projecte Wikidata. Mentre que en els primers anys els usuaris discutien sobretot vincles familiars, fonts i restriccions de propietat, amb una mirada destacada a les identitats com a binàries, en els anys següents, la discussió es va traslladar a temes més específics com ara les identitats transgènere a partir de 2019, la representació de gènere en diferents àmbits, i en general un interès més fort per la comunitat LGBTQIA+ (Metilli & Paolini, 2023).

6.4. Conclusions de la presència o absència de biaix de gènere

A partir del marc teòric revisat i de l'anàlisi i avaluació del sistema d'organització del coneixement de Wikidata, es pot concloure que:

- **Wikidata mostra més sensibilitat a la diversitat de gènere.** Si prenem el cas de la propietat de gènere P21 aporta classes relatives a gènere, com ara: agènere, dona, home, intersexual, no-binari, transgènere dona i transgènere home, entre d'altres.
- **Convindria distingir una propietat pel sexe i una pel gènere a Wikidata.** Cal revisar l'etiqueta de la propietat sexe o gènere (P21) de Wikidata que posa en una mateixa sac l'element biològic (sexe) i la construcció social del gènere, o autodefinició voluntària (gènere). La proposició disjuntiva a través de l'operador “o” (OR) associa els dos termes i, per tant, equival a la unió dels dos conjunts.
- Estudis recents han comparat el biaix de gènere de Wikidata amb la realitat, i han demostrat que **Wikidata és un fidel reflex de la realitat i que no afegeix un increment al desequilibri, tal com sí que fa Viquipèdia en els seus processos de presa de decisions** (deliberacions, votacions, etc.) pel fort component sòcio-cultural que acaba passant per damunt d'arguments més objectivables de caràcter tècnico-científic.
- **La comunitat voluntària que dona suport al projecte Wikidata intervé en les deliberacions amb argumentacions més tècniques que no pas de caràcter sociocultural,** com han estat els arguments en contra de les categories de gènere a Viquipèdia.
- **Wikidata és un *dataset* multilingüe i, per tant, cada comunitat lingüística es responsabilitza del manteniment de la traducció de les etiquetes i la presa de decisions relacionades.** Així, en el cas de les llengües que tenen flexió gramatical diferenciada per masculí i femení, l'etiquetatge de les propietats i classes relatives a persones (com ara l'ocupació) dependrà que estigui en masculí genèric o que es desdoblí per gènere a la comunitat. Però en el cas de la catalana no s'accepta el desdoblament per gènere i, per tant, **no es permet cercar per la forma femenina al cercador.** I aquest biaix és més greu en el cas de les classes perquè no totes estan disponibles en català, i, per tant, directament **no es localitzen.**
- **L'ontologia de Wikidata és una eina amb grans potencialitats tant des del punt de vista de l'adequació a sistemes de cerca orientats a l'usuari final, com des del punt de vista del seu alineament amb la perspectiva de gènere:**
 - Caràcter factual de les dades, i del seu marc de representació. **El caràcter factual de les dades ajuda a mantenir la perspectiva de gènere en el sentit que el gènere és objectivable,** es poden fonts d'informació externes i contrastades que donin directament la dada.
 - **Els responsables són conscients de les febleses que té l'ontologia, les han identificat, i estan incorporant mesures de correcció. Entre aquestes**

mesures, n'hi ha de ja incorporades, i d'altres que es projecten com a millora vinculada a la substitució del programari de fons (*backend*).

- Ja és una realitat el desenvolupament d'eines d'edició i enriquiment de dades orientades a la sistematització i consistència de les propietats, classes i etiquetes. Entre aquestes eines, destaquem els *Entity Schemas*.
- És una previsió a mig termini l'aproximació a llenguatges i tecnologies del Web semàntic, i l'adopció dels seus estàndards i tecnologies per al control de la consistència de l'ontologia, i la inferència de nou coneixement. Entre aquests estàndards, destaquem SHACL, per a la creació i validació de grafs RDF.

7. Proposta de millora del sistema d'organització del coneixement de Viquipèdia

El Moviment Wikimedia va acordar el 2017 un nou pla estratègic cap a l'horitzó 2030, que estableix l'objectiu “d'oferir coneixement com a servei” (“convertir-se en una plataforma que ofereixi el coneixement obert al món a través d'interfícies i comunitats”), tenint en compte l’“equitat de coneixement” (“enfocar el nostre esforç en el coneixement i les comunitats que han quedat al marge de les estructures de poder i privilegis [...]. Trencant les barreres socials, polítiques i tècniques que impedeixin que les persones accedeixin i contribueixin al coneixement lliure”). Aquest document estratègic, desenvolupat de manera col·laborativa, ha establert en el nucli de les seves recomanacions dos principis: el d'inclusió i el centrat en les persones (entès com a l'atenció a les necessitats de les persones). Estableix que l'objectiu per l'any 2030 és el d'eliminar la bretxa de gènere i centrar-se en la inclusió de col·lectius infrarepresentats.

La proposta de sistema d'organització que presentem en aquest document està totalment alineada amb la nova direcció estratègica. Així, per visibilitzar i donar accés al coneixement de Viquipèdia, de i sobre col·lectius minoritzats per raons de gènere (dones, persones no-binàries, intersexuals, trans homes i trans dones), es proposa una solució doble. D'una banda, una solució tècnica que implica emprar les ontologies de Wikidata com a sistema d'organització del coneixement per facilitar la cerca i recuperació de la informació a Viquipèdia, sense incrementar els biaixos existents a la realitat. I de l'altra banda, una solució de caràcter social, cultural i polític que comporta treballar amb la comunitat viquipedista a favor de l'acceptació dels estàndards internacionals reconeguts en l'àmbit dels sistemes d'organització del coneixement, i a favor de l'apropiació de dos principis, considerats estratègics pel moviment wikipedista, com són el d'inclusió per evitar la discriminació de grups minoritzats i, sobretot, el principi del servei centrat en les persones amb especial atenció a les seves necessitats informatives.

Amb la solució tècnica, en cas d'optar per les ontologies de Wikidata com a sistema d'organització del coneixement de Viquipèdia, s'aconseguiria una alineació amb els estàndards internacionals d'organització del coneixement i al mateix temps s'aconseguiria empoderar a les persones usuàries de Viquipèdia (lectores i editores) a l'hora de cercar i recuperar els continguts de l'enciclopèdia. L'empoderament rauria en el moment de la cerca i la navegació, ja que serien els usuaris els que decidirien els elements de cerca, més propers a les seves necessitats, i no com ara amb les categories, elements proposats d'acord amb la cosmovisió de les persones que les varen crear, classificar i indexar, sense permetre combinar els elements de cerca.

Entrant en detall en la proposta d'aquest document, s'apostaria per l'ús d'ontologies com a sistema d'organització de coneixement, i s'optaria per Wikidata perquè justament el propòsit original de Wikidata és l'emmagatzematge de les dades (propietats i relacions) del contingut present en els articles de Wikipedia, en qualsevol llengua. Aquesta funció, si es porta al nivell màxim, seria que Wikidata fos el sistema d'organització del coneixement de Wikipedia. És a dir,

que les categories de Wikidata es construïssin a partir de les ontologies de Wikidata. En aquest sentit, alguna categoria actual de Wikipedia ja es construeix directament des de Wikidata (“persones vives”). A la Viquipèdia ja hi ha diferents apartats vinculats a Wikidata actuant com a sistema d'organització del coneixement a través de dos exemples habituals, com són les Infotables o la secció de registres d'autoritats.

En l'avaluació feta per experts, en l'heurística, la majoria de valoracions sobre les variables d'anàlisi de Viquipèdia són baixes. En canvi, en el cas de Wikidata, n'hi ha amb marge de millora, però la comunitat de Wikidata ja les ha detectat i s'han apuntat a l'agenda processos de millora i desenvolupament d'eines per resoldre-ho. Convé destacar dues aportacions interessants com són l'entity schema i el backend, ja que comporten beneficis substancials en l'ontologia, com ara els aspectes relacionats amb l'organització de les dades relatives al sexe o gènere. Aquest aspecte demostra la potencialitat de millora. Convé també reconèixer la necessitat de revisar el nom de la propietat “sexe o gènere” que barreja en la seva etiqueta característiques biològiques i les de definició individual i construcció social.

En el cas de Wikidata, els processos de decisió per avançar en millores es decideixen en una comunitat més reduïda i amb una perspectiva més respectuosa amb els estàndards internacionals acceptats i reconeguts en l'àmbit de l'organització del coneixement. En canvi, en el cas de Viquipèdia, el biaix de gènere existent a la societat s'incrementa a partir de posicionaments contraris a la diversitat de gènere i que s'expressa amb arguments amb forta càrrega ideològica. Els arguments que apareixen en les negociacions per a la presa de decisions d'aspectes relacionats tant amb el funcionament (directrius) com en els articles (què és i què no és coneixement) no estan guiats per l'objectivitat, ni la seva aplicació és per igual en contextos similars, ni estan fonamentats en estàndards d'especificacions ja contrastats en l'àmbit professional i acadèmic.

Alhora, optar per les ontologies de Wikidata com a sistema d'organització del coneixement dels continguts de Viquipèdia empoderaria a la persona usuària (lectora o editora) per satisfer les seves necessitats informatives. Estem d'acord que el sistema de categorització de Viquipèdia (els esquemes de categories) són senzills d'entendre pels usuaris i són propers al seu vocabulari i llengua natural, però comporten desavantatges també, com els que ens ocupen en aquesta proposta, els relacionats amb la diversitat i els biaixos culturals, socials o polítics existents en la societat i els desequilibris que comporta qualsevol vocabulari controlat. Alhora, els esquemes de categories, com hem vist, són tesaures pre-coordinats que combinen conceptes, classes o termes d'un vocabulari controlat en el moment de la seva construcció, o en el moment de la indexació o classificació. Això voldria dir que pot existir una categoria de “metges catalans del sud” creada per la comunitat viquipedista i que una persona de qualsevol lloc del món hauria de poder entendre (deduir) que pot incloure “dones metges que viuen a la part del sud del principat de Catalunya”. En canvi, l'ús de les ontologies per organitzar el coneixement, s'aportaria una millor representació dels continguts de Wikidata, ja que cada propietat representa una única dimensió (atribut) d'una entitat o un conjunt d'entitats. És la persona usuària qui, en el moment de la cerca, combina els atributs que millor responen a la seva necessitat per a recuperar la o les entitats rellevants. Actualment, una categoria de la Viquipèdia com “metges catalans del sud” lliga dues dimensions diferents d'una persona: la professió i l'origen. En canvi, en una ontologia es podria triar quins elements es vol combinar (professió, any de naixement, lloc de residència o qualsevol altre de forma independent i combinada amb operadors booleans, de forma que la cerca s'acostaria molt més a la necessitat d'informació que necessita l'usuari (humà o màquina). El sistema d'organització del coneixement que planteja Wikidata és post-coordinat, de manera que empodera a les persones usuàries a través del sistema, podent combinar les característiques predefinides per l'indexador (els termes pre-coordinats). Per tant, Wikidata és l'ontologia que podria aportar organització i una millor representació del que es coneix a la Viquipèdia. De fet, la Viquipèdia ja genera categories que sorgeixen de Wikidata, com ara "persones vives".

L'ontologia de Wikidata pot ser una solució a l'aplicació de la perspectiva de gènere i a la superació de l'absència de fonamentació científico-tècnica. Amb tot, no es pot obviar la necessitat d'un canvi cultural a mig termini en el si de la comunitat, la qual hauria d'acceptar el procés de substitució per apostar per l'alineació amb els consensos internacionals en l'àmbit tecnològic i una aposta per l'accés als continguts de forma més equitativa amb la diversitat. En aquest sentit, Amical Wikimedia, ja va acceptar ser prova pilot en la junta de desembre de 2022, per incorporar Wikidata com a sistema d'organització de Viquipèdia a través d'un sistema participatiu. D'aquesta manera es podrà demostrar que és possible una passarel·la entre les ontologies de Wikidata i els continguts de Viquipèdia, i vincular així les categories (esquema de categorització) amb les propietats i classes de Wikidata. L'operativitat d'aquest procés hauria de ser participatiu i col·laboratiu, i es podria dur a terme a través de grups de treball de viquipedistes motivats per l'augment de la diversitat i voluntaris tecnòlegs.

8. Bibliografia

- Abián, David; Meroño-Peñuela, Albert; Simperl, Elena (2022). An Analysis of Content Gaps Versus User Needs in the Wikidata Knowledge Graph. In U. Sattler, A. Hogan, M. Keet, V. Presutti, J. P. A. Almeida, H. Takeda, P. Monnin, G. Pirrò, & C. d'Amato (Eds.), *The Semantic Web – ISWC 2022* (pp. 354–374). Springer International Publishing. https://doi.org/10.1007/978-3-031-19433-7_21
- Acey, Camille E.; Bouterse, Siko; Ghoshal, Sucheta; Menking, Amanda; Sengupta, Anasuya; Vrana, Adele G. (2021). Decolonizing the Internet by Decolonizing Ourselves: Challenging Epistemic Injustice through Feminist Practice. *Global Perspectives*, 2(1). <https://doi.org/10.1525/gp.2021.21268>
- Aghaebrahimian, Ahmad; Stauder, Andy; Ustaszewski, A. Michael (2022). Testing the validity of Wikipedia categories for subject matter labelling of open-domain corpus data. *Journal of Information Science*, 48(5), 686–700. <https://doi.org/10.1177/0165551520977438>
- Albuquerque, Fernando Antonio de Araújo Chacon de (2017). *Arcabouço de arquitetura da informação para ciclo de vida de projeto de vocabulário controlado: Uma aplicação em Engenharia de Software* [Fernando Antônio de Araújo Chacon de]. <https://repositorio.unb.br/handle/10482/31288>
- Alós-Moner, Adela d'; Giralt, Olga; Centelles, Miquel (2010). *Taxonomía para la web de la Junta de Andalucía. Fase 3: Aseguramiento de la calidad de la taxonomía temática. Indicadores de gestión Informe preliminar*. Junta de Andalucía.
- Amith, Muhammad; He, Zhe; Bian, Jiang; Lossio-Ventura, Juan Antonio; Tao, Cui (2018). Assessing the practice of biomedical ontology evaluation: Gaps and opportunities. *Journal of Biomedical Informatics*, 80, 1–13. <https://doi.org/10.1016/j.jbi.2018.02.010>
- Antin, Judd; Yee, Raymond; Cheshire, Coye; Nov, Oded (2011). Gender differences in Wikipedia editing. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, 11–14. <https://doi.org/10.1145/2038558.2038561>
- Bear, Julia B.; Collier, Benjamin (2016). Where are the Women in Wikipedia? Understanding the Different Psychological Experiences of Men and Women in Wikipedia. *Sex Roles*, 74(5), 254–265. <https://doi.org/10.1007/s11199-015-0573-y>
- Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina; Shmitchell, Shmargaret (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>

- Benkler, Yochai (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press. <https://www.jstor.org/stable/j.ctt1njknw>
- Black, Laura; Welser, T. DeGroot; Jocelyn, M.; Cosley, Dan (2008, May 21). "Wikipedia is Not a Democracy": Deliberation and Policy-Making in an Online Community. *Proceedings of the 58th Annual Conference of the International Communication Association*. [https://www.semanticscholar.org/paper/"Wikipedia-is-Not-a-Democracy":-Deliberation-and-in-Black-Welser/25e5fde26da12f3b140e61f6ed678e44324cbd5e](https://www.semanticscholar.org/paper/)
- Bolotnikova, Ekaterina S.; Gavrilova, Tatiana A.; Gorovoy, Vladimir A. (2011). To a method of evaluating ontologies. *Journal of Computer and Systems Sciences International*, 50(3), 448–461. <https://doi.org/10.1134/S1064230711010072>
- Booth, Andrew; Sutton, Anthea; Clowes, Mark, & Martyn-St James, Marrissa (2022). *Systematic Approaches to a Successful Literature Review* (3rd ed.). Sage Publications <https://us.sagepub.com/en-us/nam/systematic-approaches-to-a-successful-literature-review/book270933>
- Brückner, Hannah; Adams, Julia (2015). Wikipedia, sociology, and the promise and pitfalls of Big Data. *Big Data and Society*, 2(2). Buchem, Ilona; Kloppenburg, Julia (2013). *Gender – Diversität – Wikipedia*. Beuth Hochschule für Technik Berlin. https://www.bht-berlin.de/fileadmin/oe/gutz/Sonstige_Veroeffentlichungen/Arbeitspapier_Gender-Diversity-Wikipedia.pdf
- Buolamwini, Joy; Gebru, Timnit (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Calvo Iglesias, Encina (2020). Preparing Biographies of STEM Women in the Wikipedia Format, a Teaching Experience. *Revista Iberoamericana de Tecnologías Del Aprendizaje*, 15(3), 211–214. <https://doi.org/10.1109/RITA.2020.3008144>
- Castells, Manuel (2008). The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance. *The ANNALS of the American Academy of Political and Social Science*, 616(1), 78–93. <https://doi.org/10.1177/0002716207311877>
- Clua, Anna; Ferran-Ferrer, Núria; Terren, Ludovic (2018). El impacto de los jóvenes en la esfera pública: La disolución del Consejo de la Juventud de España en la prensa y en Twitter. *Comunicar: Revista Científica de Comunicación y Educación*, 26(55), 49–57. <https://doi.org/10.3916/C55-2018-05>
- Cohen, Jacob (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Collier, Benjamin; Bear, Julia (2012). Conflict, criticism, or confidence: An empirical examination of the gender gap in wikipedia contributions. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 383–392. <https://doi.org/10.1145/2145204.2145265>
- da Costa, Tiago Vinícius Remígio; Cavalcante, Everton; Batista, Thais (2022). Big Data Software Architectures: An Updated Review. In O. Gervasi, B. Murgante, E. M. T. Hendrix, D. Taniar, & B. O. Apduhan (Eds.), *Computational Science and Its Applications – ICCSA 2022* (pp. 477–493). Springer International Publishing. https://doi.org/10.1007/978-3-031-10522-7_33
- van Dijk, Jan (2012). *The Network Society* (3rd ed.). SAGE Publications Ltd. <https://us.sagepub.com/en-us/nam/the-network-society/book268672>
- Eckert, Stine; Steiner, Linda (2013). (Re)triggering Backlash: Responses to News About Wikipedia's Gender Gap. *Journal of Communication Inquiry*, 37(4), 284–303. <https://doi.org/10.1177/0196859913505618>

- Erleben, Fredo; Günther, Michael; Kröttsch, Markus; Mendez, Julian; Vrandečić, Denny (2014). Introducing Wikidata to the Linked Data Web. In P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, & C. Goble (Eds.), *The Semantic Web – ISWC 2014* (pp. 50–65). Springer International Publishing. https://doi.org/10.1007/978-3-319-11964-9_4
- Evans, Sian; Mabey, Jacqueline; Mandiberg, Michael (2015). Editing for Equality: The Outcomes of the Art+Feminism Wikipedia Edit-a-thons. *Art Documentation*, 34(2), 194–203. <https://doi.org/10.1086/683380>
- Farzan, Rosta; Savage, Saiph; Saviaga, Claudia Flores (2016). Bring on Board New Enthusiasts! A Case Study of Impact of Wikipedia Art + Feminism Edit-A-Thon Events on Newcomers. In E. Spiro & Y.-Y. Ahn (Eds.), *Social Informatics* (pp. 24–40). Springer International Publishing. https://doi.org/10.1007/978-3-319-47880-7_2
- Ferran-Ferrer, Núria; Castellanos-Pineda, Patricia; Minguillón, Julià; Meneses, Julio (2021). The gender gap on the Spanish Wikipedia: Listening to the voices of women editors. *Profesional de La Información*, 30(5), Article 5. <https://doi.org/10.3145/epi.2021.sep.16>
- Ferran-Ferrer, Núria; Miquel-Ribé, Marc; Meneses, Julio; Minguillón, Julià (2022). The Gender Perspective in Wikipedia: A Content and Participation Challenge. *Companion Proceedings of the Web Conference 2022*, 1319–1323. <https://doi.org/10.1145/3487553.3524937>
- Ford, Heather; Wajcman, Judy (2017). "Anyone can edit", not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, 47(4), 511–527. <https://doi.org/10.1177/0306312717692172>
- Gardner, Sue (2011, February 20). Nine Reasons Women Don't Edit Wikipedia (in their own words). *Sue Gardner's Blog*. <https://suegardner.org/2011/02/19/nine-reasons-why-women-dont-edit-wikipedia-in-their-own-words/>
- Gerguis, Michael Naim; Salama, Cherif; El-Kharashi, M. Watheq (2017). WikiTrends: Unstructured Wikipedia-Based Text Analytics Framework. In F. Frasincar, A. Ittoo, L. M. Nguyen, & E. Métails (Eds.), *Natural Language Processing and Information Systems* (pp. 45–57). Springer International Publishing. https://doi.org/10.1007/978-3-319-59569-6_6
- Gil Leiva, Isidoro (2008). *Manual de indización: Teoría y práctica*. <https://dialnet.unirioja.es/servlet/libro?codigo=609726>
- Gleißner, Lea-Kathrin; Bui, Magdalena; Kühn, Fey; Nenninger, Amelie (2021). *Designing an ethical technology project with the help of Data Feminism*. Gesellschaft für Informatik, Bonn. <https://dl.gi.de/handle/20.500.12116/37775>
- Graells-Garrido, Eduardo; Lalmas, Mounia; Menczer, Filippo (2015). First Women, Second Sex: Gender Bias in Wikipedia. *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, 165–174. <https://doi.org/10.1145/2700171.2791036>
- Grant, Maria J.; Booth, Andrew (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Gruwell, Leigh (2015). Wikipedia's Politics of Exclusion: Gender, Epistemology, and Feminist Rhetorical (In)action. *Computers and Composition*, 37, 117–131. <https://doi.org/10.1016/j.compcom.2015.06.009>
- Hevia, Alejandro González; Facundo, Guillermo; Azcona, Emilio Rubiera; Gayo, José Emilio Labra (2020). Automatic synchronization of RDF graphs representing ontologies and Wikibase instances. *Proceedings of the 1st Wikidata Workshop (Wikidata 2020)*. <https://ceur-ws.org/Vol-2773/paper-07.pdf>

- Hinnosaar, Marit (2019). Gender inequality in new media: Evidence from Wikipedia. *Journal of Economic Behavior & Organization*, 163, 262–276. <https://doi.org/10.1016/j.jebo.2019.04.020>
- Hollink, Laura; van Aggelen, Astrid; van Ossenbruggen, Jacco (2018). Using the Web of Data to Study Gender Differences in Online Knowledge Sources: The Case of the European Parliament. *Proceedings of the 10th ACM Conference on Web Science*, 381–385. <https://doi.org/10.1145/3201064.3201108>
- Hood, Nina; Littlejohn, Allison (2018). Hacking History: Redressing Gender Inequities on Wikipedia Through an Editathon. *The International Review of Research in Open and Distributed Learning*, 19(5). <https://doi.org/10.19173/irrodl.v19i5.3549>
- Hwang, Sungsoo (2008). Utilizing Qualitative Data Analysis Software: A Review of Atlas.ti. *Social Science Computer Review*, 26(4), 519–527. <https://doi.org/10.1177/0894439307312485>
- Kaplan, Angelika; Kühn, Thomas; Hahner, Sebastian; Benkler, Niko; Keim, Jan; Fuchß, Dominik; Corallo, Sohie; Heinrich, Robert (2022). Introducing an Evaluation Method for Taxonomies. *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering*, 311–316. <https://doi.org/10.1145/3530019.3535305>
- Kaffee, Lucie-Aimée; Piscopo, Alessandro; Vougiouklis, Pavlos; Simperl, Elena; Carr, Leslie; Pintscher, Lydia (2017). A Glimpse into Babel: An Analysis of Multilinguality in Wikidata. *Proceedings of the 13th International Symposium on Open Collaboration*, 1–5. <https://doi.org/10.1145/3125433.3125465>
- Karczewska, Anna; Kukowska, Katarzyna (2021). Cultural dimension of femininity: Masculinity in virtual organizing knowledge sharing. *Proceedings of the European Conference on Knowledge Management*, 414–422. <https://www.proquest.com/docview/2581071766>
- Kessler, Suzanne J.; McKenna, Wendy (1985). *Gender: An Ethnomethodological Approach*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/G/bo3629888.html>
- Keyes, Os (2018). The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 88:1-88:22. <https://doi.org/10.1145/3274357>
- Klein, Maximilian; Gupta, Harsh; Rai, Vivek; Konieczny, Piotr; Zhu, Haiyi (2016). Monitoring the gender gap with Wikidata human gender indicators. *Proceedings of the 12th International Symposium on Open Collaboration*. Scopus. <https://doi.org/10.1145/2957792.2957798>
- Kless, Daniel; Milton, Simon (2010). Towards Quality Measures for Evaluating Thesauri. In S. Sánchez-Alonso & I. N. Athanasiadis (Eds.), *Metadata and Semantic Research* (pp. 312–319). Springer. https://doi.org/10.1007/978-3-642-16552-8_28
- Konieczny, Piotr; Klein, Maximilian (2018). Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator. *New Media & Society*, 20(12), 4608–4633. <https://doi.org/10.1177/1461444818779080>
- Lancaster, Frederick Wilfrid (2002). *El control del vocabulario en la recuperación de información* (2ª ed.). Universidad de Valencia <https://dialnet.unirioja.es/servlet/libro?codigo=729313>
- Leonard, Victoria; Bond, Sarah E. (2019). Advancing feminism online: Online tools, visibility, and women in classics. *Studies in Late Antiquity*, 3(1), 4–16. Scopus. <https://doi.org/10.1525/sla.2019.3.1.4>
- Macharia, Sarah (2020). Global Media Monitoring Project (GMMP). In *The International Encyclopedia of Gender, Media, and Communication* (pp. 1–6). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119429128.iegmc074>

- Matud, M. Pilar; Espinosa, Inmaculada; Rodríguez Wangüemert, Carmen (2021). Women and men portrayal on television news: A study of Spanish television newscast. *Feminist Media Studies*, 21(2), 298–314. <https://doi.org/10.1080/14680777.2019.1681489>
- Mazzocchi, Fulvio (2018). Knowledge Organization System (KOS): An Introductory Critical Account. *Knowledge Organization*, 45(1). <https://doi.org/10.5771/0943-7444-2018-1-54>
- Mehrabi, Ninareh; Morstatter, Fred; Saxena, Nripsuta; Lerman, Kristina; Galstyan, Aram (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6), 115:1-115:35. <https://doi.org/10.1145/3457607>
- Menking, Amanda; Erickson, Ingrid (2015). The Heart Work of Wikipedia: Gendered, Emotional Labor in the World’s Largest Online Encyclopedia. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 207–210. <https://doi.org/10.1145/2702123.2702514>
- Menking, Amanda; Rosenberg, Jon (2021). WP:NOT, WP:NPOV, and Other Stories Wikipedia Tells Us: A Feminist Critique of Wikipedia’s Epistemology. *Science Technology and Human Values*, 46(3), 455–479. Scopus. <https://doi.org/10.1177/0162243920924783>
- Metilli, Daniele; Paolini, Chiara (2023). *Non-binary Gender Representation in Wikidata*. <https://wgedi.com/chapter.pdf>
- Minguillón, Julià; Lerga, Maura; Aibar, Eduard; Lladós-Masllorens, Josep; Meseguer-Artola, Antoni (2017). Semi-automatic generation of a corpus of Wikipedia articles on science and technology. *Profesional de la Información*, 26(5), Article 5. <https://doi.org/10.3145/epi.2017.sep.20>
- Minguillón, Julià; Meneses, Julio; Aibar, Eduard; Ferran-Ferrer, Núria; Fàbregues, Sergi (2021). Exploring the gender gap in the Spanish Wikipedia: Differences in engagement and editing practices. *PLoS ONE*, 16(2 February). Scopus. <https://doi.org/10.1371/journal.pone.0246702>
- Miquel-Ribe, Marc; Laniado, Daniel (2021). The Wikipedia Diversity Observatory: Helping communities to bridge content gaps through interactive interfaces. *Journal of Internet Services and Applications*, 12(1). <https://doi.org/10.1186/s13174-021-00141-y>
- Mora-Cantalops, M., Sánchez-Alonso, S., & García-Barriocanal, E. (2019). A systematic literature review on Wikidata. *Data Technologies and Applications*, 53(3), 250–268. <https://doi.org/10.1108/DTA-12-2018-0110>
- Morales-Vargas, Alejandro; Pedraza-Jimenez, Rafael; Codina, Lluís (2023). Website quality evaluation: A model for developing comprehensive assessment instruments based on key quality factors. *Journal of Documentation*, 79(7), 95–114. <https://doi.org/10.1108/JD-11-2022-0246>
- Morgan, Jonathan T; Bouterse, Siko; Walls, Heather; Stierch, Sarah (2013). Tea and sympathy: Crafting positive new user experiences on wikipedia. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 839–848. <https://doi.org/10.1145/2441776.2441871>
- Morville, Peter (2005). Experience design unplugged. *ACM SIGGRAPH 2005 Web Program*, 10-es. <https://doi.org/10.1145/1187335.1187347>
- Morville, Peter; Rosenfeld, Louis (2006). *Information Architecture for the World Wide Web* (3rd ed.). <https://www.oreilly.com/library/view/information-architecture-for/0596527349/>
- Park, Diana E.; Bridges, Laurie M. (2022). Meet Students Where They Are: Centering Wikipedia in the Classroom. *Communications in Information Literacy*, 16(1), 4–23. Scopus. <https://doi.org/10.15760/comminfolit.2022.16.1.2>
- Pasca, Marius (2018). Finding Needles in an Encyclopedic Haystack: Detecting Classes Among Wikipedia Articles. *Proceedings of the 2018 World Wide Web Conference*, 1267–1276. <https://doi.org/10.1145/3178876.3186025>

- Piscopo, Alessandro; Phethean, Chris; Simperl, Elena (2017). What Makes a Good Collaborative Knowledge Graph: Group Composition and Quality in Wikidata. In G. L. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), *Social Informatics* (pp. 305–322). Springer International Publishing. https://doi.org/10.1007/978-3-319-67217-5_19
- Piscopo, Alessandro; Simperl, Elena (2018). Who Models the World? Collaborative Ontology Creation and User Roles in Wikidata. *Proceedings of the ACM on Human-Computer Interaction*, 2 (CSCW), 141:1-141:18. <https://doi.org/10.1145/3274410>
- Preece, Jennifer; Sharp, Helen; Rogers, Yvonne (2019). *Interaction Design: Beyond Human-Computer Interaction* (5th ed.). Indianapolis: John Wiley & Sons.
- Quintarelli, Elisa. (2005). Folksonomies: Power to the people. ISKO Italy-UniMIB meeting: Milà. Juny.
- Ramírez-Ordóñez, David, & Ferran-Ferrer, Núria (2023). Endurance against oblivion: The case of the Articles for Deletion with gender perspective in Wikipedia. *Proceedings of the Wiki Workshop (10th Edition)*. https://wikiworkshop.org/2023/papers/WikiWorkshop2023_paper_34.pdf
- Ramírez Plascencia, David (2021). Medios digitales y el empoderamiento de las mujeres en América Latina: El caso de las editadoras. *Journal of Iberian and Latin American Research*, 27(3), 459–471. <https://doi.org/10.1080/13260219.2021.2030282>
- Rodó-Zárate, Maria (2021). Interseccionalitat. Desigualtats, llocs i emocions. Manresa: Tigre de Paper. <https://www.tigredpaper.cat/ca/catalogue/interseccionalitat>
- Sauro, Jeff; Lewis, James R. (2016). *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-384968-7.00013-8>
- Schneider, Jodi; Samp, Krystian; Passant, Alexandre; Decker, Stefan (2013). Arguments about deletion: How experience improves the acceptability of arguments in ad-hoc online task groups. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 1069–1080. <https://doi.org/10.1145/2441776.2441897>
- Stock, Wolfgang G. (2015). *Informetric Analyses of Knowledge Organization Systems (KOSs)* (arXiv:1505.03671). arXiv. <https://doi.org/10.48550/arXiv.1505.03671>
- Souza, Renato R.; Tudhope, Douglas; Almeida, Mauricio Barcellos (2012). Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. *Knowledge Organization*, 39(3), 179–192. <https://doi.org/10.5771/0943-7444-2012-3-179>
- Tkacz, Nathaniel (2014). *Wikipedia and the Politics of Openness*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/W/bo19085555.html>
- Tripodi, Francesca (2023). Ms. Categorized: Gender, notability, and inequality on Wikipedia. *New Media & Society*, 25(7), 1687–1707. <https://doi.org/10.1177/14614448211023772>
- Vander Wal, Thomas (2007). *Folksonomy: Vanderwal.net*. Folksonomy. <https://vanderwal.net/folksonomy.html>
- Veale, Michael; Binns, Reuben (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2), 2053951717743530. <https://doi.org/10.1177/2053951717743530>
- Vrandečić, Denny; Krötzsch, Markus (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78–85. <https://doi.org/10.1145/2629489>
- Vrandečić, Denny; Pintscher, Lydia; Krötzsch, Markus (2023). Wikidata: The Making Of. *Companion Proceedings of the ACM Web Conference 2023*, 615–624. <https://doi.org/10.1145/3543873.3585579>

- Wagner, Claudia; Garcia, David; Jadidi, Hohsen; Strohmaier, Markus (2015). It's a Man's Wikipedia? : Assessing Gender Inequality in an Online Encyclopedia. *Proceedings of the 9th International AAAI Conference on Web and Social Media*, 454–463. <https://doi.org/10.1609/icwsm.v9i1.14628>
- Wagner, Claudia; Graells-Garrido, Eduardo; Garcia, David; Menczer, Filippo (2016). Women through the glass ceiling: Gender asymmetries in Wikipedia. *EPJ Data Science*, 5(1), Article 1. <https://doi.org/10.1140/epjds/s13688-016-0066-4>
- Weathington, Katy; Brubaker, Jed R. (2023). Queer Identities, Normative Databases: Challenges to Capturing Queerness On Wikidata. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 84:1-84:26. <https://doi.org/10.1145/3579517>
- Wilson, R. Shyama I.; Goonetillake, Jeevani S.; Ginige, Athula; Walisadeera, Anusha Indika (2022). Ontology Quality Evaluation Methodology. In O. Gervasi, B. Murgante, E. M. T. Hendrix, D. Taniar, & B. O. Apduhan (Eds.), *Computational Science and Its Applications – ICCSA 2022* (pp. 509–528). Springer International Publishing. https://doi.org/10.1007/978-3-031-10522-7_35
- Worku, Zena; Bipat, Taryn; McDonald, David W.; Zachry, Mark (2020). Exploring Systematic Bias through Article Deletions on Wikipedia from a Behavioral Perspective. *Proceedings of the 16th International Symposium on Open Collaboration*, 1–22. <https://doi.org/10.1145/3412569.3412573>
- Yedid, Nadina (2013). Introducción a las Folksonomías: Definición, Características y Diferencias con los Modelos Tradicionales de Indización. *Información, cultura y sociedad*, 29, Article 29. <https://doi.org/10.34096/ics.i29.673>
- Yu, Amy Zhao; Ronen, Shahar; Hu, Kevin; Lu, Tiffany; Hidalgo, César A. (2016). Pantheon 1.0, a manually verified dataset of globally famous biographies. *Scientific Data*, 3(1), Article 1. <https://doi.org/10.1038/sdata.2015.75>
- Zeng, Marcia Lei (2008). Knowledge Organization Systems (KOS). *Knowledge Organization*, 35(2–3), 160–182. <https://doi.org/10.5771/0943-7444-2008-2-3-160>
- Zhang, Charles Chuankai; Terveen, Loren (2021). Quantifying the gap: A case study of wikidata gender disparities. *Proceedings of the 17th International Symposium on Open Collaboration*, 1–12. Scopus. <https://doi.org/10.1145/3479986.3479992>
- Zheng, Xiang; Chen, Jiajing; Yan, Erjia; Ni, Chaoqun (2022). Gender and country biases in Wikipedia citations to scholarly publications. *Journal of the Association for Information Science and Technology*. Scopus. <https://doi.org/10.1002/asi.24723>

9. Agraïments

A totes les 20 persones participants dels testos fets amb persones usuàries (lectores i editores) de Viquipèdia i Wikidata, pel seu temps en els experiments i ajut. A la comunitat voluntària de Viquipèdia, membres d'Amical i membres de la junta d'Amical Wikimedia.

A Viquidones UPF, i en especial a l'Ester Bonet, Pilar Salvatella, Meritxell Casadesús, i Núria Ribas, en concret, a la iniciativa de Dones amb Categoria, i a la Rosa Cerarols directora de la Unitat d'Igualtat UPF per elevar la preocupació de la bretxa de gènere a les categories de Viquipèdia a la Xarxa Vives d'Universitats (XUV).

Agraïments pel finançament del projecte Women and Wikipedia (W&W) (REF Reference Grant PID2020-116936RA-I00) del Ministerio de Ciencia e Innovación (MCIN) i la Agencia de Investigación de España (AEI).

A la Xarxa Vives d'Universitats, per l'encàrrec i suport.

A tot l'equip del projecte W&W per la feina feta.

10. Autories

Dra. Núria Ferran-Ferrer, professora agregada a la Facultat d'Informació i Mitjans Audiovisuals de la Universitat de Barcelona (UB) des del 2021. Coordinadora del Programa de Doctorat en Informació i Comunicació i Delegada del rector per la Direcció de la Unitat d'Igualtat. Té el doctorat Europeu per la Universitat de Barcelona (UB, 2010), llicenciada en Periodisme (UAB, 1998), en Documentació (UOC, 2003) i màster en Societat de la Informació (IN3-UOC, 2005). Ha estat professora associada en diferents universitats: UAB, UB i UPF i ha fet estades a la Universitat de Sheffield i a la Universitat de Tallin. En l'àmbit de la recerca, va començar a treballar en continguts oberts i cultura col·laborativa el 2006 en el projecte europeu OLCOS. És investigadora principal del projecte Women and Wikipedia (W&W, Reference Grant PID2020-116936RA-I00) que estudia el biaix de gènere de les dones a la Wikipedia. De forma permanent intenta col·laborar amb associacions i entitats culturals amb qui comparteix aspiracions per realitzar canvis socials. De fet, entén la recerca com una eina al servei de la societat i amb impacte sobre les seves necessitats.

Miquel Centelles, professor de la Facultat d'Informació i Mitjans Audiovisuals de la Universitat de Barcelona (UB). Expert en tecnologies semàntiques aplicades a la gestió de la informació i el coneixement. És el coordinador del màster en Humanitats Digitals on participen cinc facultats de la UB. Ha col·laborat en diferents projectes de recerca com el projecte Archiver per la preservació digital (Archiver TENDER - European Union) de dades de recerca i també en el disseny de la metodologia Sapiens per part d'elBullifoundation, que dirigeix Ferran Adrià. Ha fet estades internacionals com la més recent a la Universitat d'Illinois. És investigador del projecte Women and Wikipedia.

Yessica Macià, estudiant de doctorat del programa Informació i Comunicació de la Facultat d'Informació i Mitjans Audiovisuals de la Universitat de Barcelona (UB) des del 2023. És investigadora del projecte Women and Wikipedia.

Dr. Juan-José Boté, professor de la Facultat d'Informació i Mitjans Audiovisuals de la Universitat de Barcelona (UB). Doctor en Filosofia (Lingüística i Ciències de la Informació) per la Universität Hildesheim (2022) i doctor en Informació i Documentació en la Societat del Coneixement per la Universitat de Barcelona (2013). És membre de la comissió d'igualtat i de la comissió de qualitat de la facultat. És membre de l'Institut Interuniversitari d'Estudis de Dones i Gènere (IIEDG). Ha estat professor associat a diferents universitats: UAB, VIU, UOC. Ha participat en diferents projectes de recerca com Ciencia abierta en España: una aproximación global para evaluar su grado de implementación (RTI2018-094360-B-I00) o en el projecte Archiver per la preservació digital (Archiver TENDER - European Union) de dades de recerca. És investigador del projecte Women and Wikipedia.

Dr. Julià Minguillón, professor agregat dels Estudis d'Informàtica, Multimèdia i Telecomunicacions de la UOC des de 2001. Doctorat en Enginyeria Informàtica per la UAB l'any 2002. És investigador del projecte Women and Wikipedia.

11. Glossari

Bolcat de base de dades (en anglès, database dump): Procès i resultat d'obtenir el registre de l'estructura de la taula i/o les dades d'una base de dades. Normalment té la forma d'una llista de sentències SQL.

Categoria: Concepte representat en un sistema d'organització del coneixement. Tot i que és diferent a la designació del concepte, o terme, sovint s'utilitza la forma categoria per al conjunt dels conceptes i els termes.

Classificació: Activitat que consisteix en l'agrupament de components semblants o coses relacionades; separant les coses diferents o no relacionades; i ordenant els grups resultants en una seqüència lògica i útil. El resultat d'aquesta activitat.

Clic: Cada una de les pulsacions d'un botó o un enllaç realitzades per una persona usuària de forma intencionada. És una mesura utilitzada en els tests de persones usuàries en els sistemes de cerca de Viquièdia i de Wikidata.

Completesa: Indicador que denota la fracció de termes complets sobre tots els objectes estudiats.

Esquema de categories: Tipus de sistema d'organització del coneixement format per conceptes relacionats entre si mitjançant una relació semàntica no taxonòmica.

Etiquetatge social (en anglès, *social tagging*): Procés d'assignació d'etiquetes a continguts informatius, de forma col·laborativa, i sense regles estrictes pel que fa a la seva forma i estructura, amb l'objectiu de facilitar la cerca dels continguts, i compartir interessos.

Folksonomia (en anglès, *folksonomy*): Sistema de classificació col·laborativa i espontània de continguts d'Internet, basat en l'atribució de mots clau triats lliurement pels usuaris no especialistes, que afavoreix compartir els recursos i permet millorar la cerca d'informació.

Fons (en anglès, *backend*): Part d'un sistema informàtic, integrada per programari, bases de dades, processadors i altres components, que processa i emmagatzema les dades requerides pel front i respon a les seves sol·licituds.

Generalitat: Indicador d'avaluació de les taxonomies que mesura si és prou general i prou específica. Es vincula a la Laconicitat i la Lucidesa.

Graf dirigit: Conjunt d'objectes, normalment anomenats "nodes", connectats entre si per línies, normalment només anomenades arestes, que es dirigeixen d'un node a un altre.

Inferència: En el context del Web semàntic, procés mitjançant el qual s'afegeixen dades noves a un conjunt de dades, creades a partir de les dades existents.

Intent: mesura pels tests a Wikidata. Cada una de les vegades que la persona usuària clica el botó de "executa la consulta" en el cercador de QueryBuilder de Wikidata. Aquest cercador no funciona amb un recorregut a la navegació de la interfície, sinó que crea i executa la consulta que la persona usuària fa cada moment.

Justificació: Criteri sobre el qual es fonamenta la representació d'un concepte en un esquema de categories, o per a la selecció d'un dels termes que designen un mateix concepte com a terme preferent. La norma ANSI/NISO Z39.19-2005 (R2010) identifica tres criteris de justificació diferents: freqüència d'aparició en la bibliografia (*literary warrant*), freqüència de cerca del concepte per part dels usuaris (*user warrant*), i característiques i context de l'organització (*organizational warrant*).

Laconicitat: Indicador relatiu a fracció de termes lacònics entre tots els objectes que s'estudien. Mesura la possibilitat de les etiquetes a ser expressades amb el menor emplenament de paraules.

Lucidesa: Criteri que determina la fracció de les classes lúcides, que, al seu torn, s'han d'aproximar a la categoria genèrica.

Ortogonalitat: Indicador que mesura la superposició i els límits dels termes dins d'una categoria.

Polijerarquia: Una estructura de vocabulari controlada en la qual alguns termes pertanyen a més d'una jerarquia. Per exemple, rosa podria ser un terme més restringit tant a les flors com a les plantes perennes en un vocabulari d'horticultura.

Precoordinació: Combinació de conceptes, classes o termes d'un vocabulari controlat en el moment de la seva construcció o en el moment d'usar-lo a la indexació o classificació.

Probabilitat de retorn: Quocient entre la mitjana de retorns, és a dir, quan la persona usuària ha de tornar a la pàgina anterior per a corregir l'error de la cerca i/o navegació.

RDF. Vegeu: Resource Description Framework.

Recorregut cognitiu (en anglès, *cognitive walktrough*): Tècnica en què un expert construeix escenaris amb les tasques que la persona usuària ha de fer, i les executa assumint el rol de l'usuari. Durant l'execució d'aquesta tècnica, s'analitzen les tasques que la persona usuària ha de dur a terme i se simula el procés de resolució dels problemes de cada etapa del procés d'interacció.

Repositori RDF (en anglès, *triplestore*): Base de dades especialment dissenyada per a l'emmagatzematge i la recuperació de declaracions o tripletes RDF.

Resource Description Framework: Model de dades, estandarditzat pel World Wide Web Consortium (W3C), i usat per definir dades en el web semàntic i, en general, en les aplicacions que requereixen un estàndard per intercanviar dades. La seva abreviatura és RDF.

SHACL: Vegeu *Shapes Constraint Language*.

Shapes Constraint Language: Llenguatge per a la descripció i validació de grafs RDF. És un estàndard del World Wide Web Consortium en el context del Web semàntic. La seva abreviació és SHACL.

Simple Knowledge Organization System: Vocabulari RDF per representar sistemes d'organització del coneixement com ara taxonomies o esquemes de categorització. És un estàndard del World Wide Web Consortium en el context del Web semàntic. La seva abreviació és SKOS.

Sistema d'organització del coneixement (en anglès, *knowledge organization system*): Sistema que modela l'estructura semàntica subjacent d'un domini, i proporcionar semàntica, navegació i traducció mitjançant etiquetes, definicions, relacions i propietats dels conceptes. La seva abreviació és SOC.

SKOS: Vegeu: *Simple Knowledge Organization System*.

SOC: Vegeu Sistema d'organització del coneixement.

Sonoritat: Indicador relatiu a fa referència a la fracció de les classes de so a la taxonomia.

SPARQL: Vegeu SPARQL Protocol and RDF Query Language.

SPARQL Protocol and RDF Query Language: Llenguatge estandarditzat per a la consulta de bases de dades, capaç de recuperar i manipular dades en format RDF. És un estàndard del World Wide Web Consortium en el context del Web semàntic. La seva abreviació és SPARQL.

Subclasse directa: En les ontologies, classe vinculada amb la seva superclasse de forma explícita, mitjançant una propietat de subclassificació.

Subclasse indirecta: En les ontologies, classe vinculada amb la seva superclasse de forma implícita, a través de l'execució de la transitivitat amb què es caracteritza la propietat de subclassificació, mitjançant mecanismes d'Inferència.

Taxonomia: Tipus de sistema d'organització del coneixement estructurat mitjançant divisions d'elements en grups o categories ordenats en funció de característiques particulars.

Viquipèdia: l'enciclopèdia lliure que tothom pot editar en llengua catalana <https://ca.wikipedia.org/>. Versió catalana de Wikipedia.

Web semàntic: Projecte que té com a objectiu crear un medi universal per a l'intercanvi d'informació significativa (semàntica), d'una forma comprensible per a les màquines, del contingut dels documents del web. És impulsat pel World Wide Web Consortium.

Wikipedia: enciclopèdia col·laborativa, oberta i digital en gairebé 300 idiomes i d'abast mundial. L'ús de la paraula Wikipedia en aquest informe es reserva per a la referència a l'enciclopèdia en el seu conjunt, incloent-hi totes les edicions lingüístiques.

Llistat de taules, imatges i figures

Llistat de Taules

Taula 1. Síntesi dels objectius, materials i mètodes del projecte. pàg. 11

Taula 2. Adaptat del marc SALSA per l'scoping review de Wikipedia i gènere. pàg. 13

Taula 3. Adaptat del marc SALSA per l'scoping review dels sistemes d'organització del coneixement a Wikipedia i Wikidata. pàg. 14

Taula 4. Persones expertes entrevistades. pàg. 15

- Taula 5. Professions feminitzades a Catalunya. pàg. 16
- Taula 6. Distribució per característiques essencials de la mostra emprada pels tests amb 20 persones usuàries. pàg. 17
- Taula 7. Experiments per al test amb les persones usuàries sobre les categories de Viquipèdia i les ontologies de Wikidata. pàg. 18
- Taula 8. Indicadors utilitzats per l'avaluació dels sistemes de navegació i cerca de Viquipèdia i Wikidata. pàg. 24
- Taula 9. Indicadors utilitzats en l'avaluació heurística de l'esquema de categories de Viquipèdia. pàg. 27
- Taula 10. Llibre de codis d'anàlisi de les deliberacions de les categories d'identitats de gènere "dones" i "persones no binàries" a la Viquipèdia. pàg. 30
- Taula 11. Objectiu general de la proposta. pàg. 40
- Taula 12. Objectius específics de la proposta. pàg. 40
- Taula 13. Resum de la caracterització en 2 entorns de les folksonomies. pàg. 43
- Taula 14. Eficiència de cerca i navegació a les diferents opcions de Viquipèdia en comparació a la probabilitat de retorn. pàg. 54
- Taula 15. Valoració mitjana dels cercadors. pàg. 57
- Taula 16. Nombre de persones usuàries que satisfan la necessitat d'informació plantejada en cada tasca. pàg. 58
- Taula 17. Satisfacció amb sistemes de cerca i navegació de la Viquipèdia per cada una de les tasques plantejades. pàg. 59
- Taula 18. Posicionament d'arguments en les deliberacions de l'acceptació o rebuig de les categories "dones" (D1 i D2) i "persones no-binàries (D3). pàg. 68
- Taula 19. Votació sobre la creació de la categoria de "dones" a la Viquipèdia catalana. pàg. 68
- Taula 20. Índex d'eficiència del cercador de Wikidata 'QueryBuilder' (tasca 9). pàg. 87
- Taula 21. Índex d'eficiència i probabilitat d'error a la navegació per les llistes de propietats de Wikidata. pàg. 88
- Taula 22. Nombre de persones que satisfan la necessitat informativa amb el cercador de Wikidata pàg. 89

Llistat d'imatges

- Imatge 1. Portada de la Viquipèdia. pàg. 19
- Imatge 2. Cercador general. pàg. 20
- Imatge 3. Cercador de gènere. pàg. 20
- Imatge 4. Cercador intern. pàg. 21
- Imatge 5. Cercador interseccional. pàg. 22
- Imatge 6. Arbre de categories de Viquipèdia. pàg. 21
- Imatge 7. Índex de categories. pàg. 21
- Imatge 8. Cercador de Wikidata amb QueryBuilder. pàg. 22
- Imatge 9. Llistat de propietats de Wikidata. pàg. 22
- Imatge 10. Representació gràfica d'una unitat mínima d'un graf.
- Imatge 10. Element "Bel Olid Báez Q16176248" a Wikidata. pàg. 72
- Imatge 11. Propietat 21 de sexe o gènere per l'element "Bel Olid Báez (Q16176248)" a Wikidata. pàg. 73
- Imatge 12. Nom de naixement per l'element "Bel Olid Báez (Q16176248)" a Wikidata. pàg. 73
- Imatge 13. Restricció un-de de la propietat sexe o gènere, pàg. 75

Llistat de figures

- Figura 1. Indicador Estabilitat. pàg. 62
- Figura 2. Indicador: Estabilitat. Categories i Acumulat. pàg. 63
- Figura 3. Nombre de subcategories i pàgines per cada categoria principal. pàg. 66
- Figura 4. Indicador: Estabilitat. Categories i Acumulat. pàg. 66
- Figura 5. Exemple de cicle o bucle de la propietat subclasse de (P279). pàg. 92

Figura 6. Exemple de generalització redundat (1). pàg. 92

Figura 7. Exemple de generalització redundat (2). pàg. 93

Annexos

Annex A. Guió entrevistes expertes

1. Questions about realization and performance of services for data sharing and query.

As it's profusely spread, Wikidata is discontinuing Blazegraph as a triplestore and backend for Wikidata Query Service due to a conjunctural cause (the purchase of this software by Amazon, which will cancel short- or long-term maintenance/support strategy), and to other causes of technological limitation.

- It's a single-server implementation; it suffers performance problems at the current database size of ~14B triples.
 - Query performance is based on synchronous response and 'older' technology, so as to experience timeouts and general instability.
1. We'd like to have information about the current features of triplestore and backend for Wikidata Query Service with Blazegraph, and improvements you envisage with the new configuration.
 1. Current configuration of the backend to face queries based on the following features.
 - What's the query answering timeout and maximum term before abortion?
 - What's the number of concurrent queries per client?
 - What's the answering time average?
 - What are the features concerning the limited amount of processing time and query errors?
 - What are the limits on result size?
 2. Based on the features, is it strong enough to face the increasing query traffic? Do you consider that these features would be maintainable if the service is provided for searching Wikipedia content?
 3. Is it possible to access the core load and performance figures? (i.e. Wikidata Query Service dashboard?) Is this the source?
<https://grafana.wikimedia.org/d/000000489/wikidata-query-service?orgId=1&refresh=1m&viewPanel=7>
https://wikidata.org/wiki/Wikidata:Data_access
 2. Improvements with the new WDQS backend.
 1. What are the specific goals to reach concerning scalability, federated query, query timeout and resource recovery?
 2. Considering the "Technical and community criteria assessments"
https://wikidata.org/wiki/Wikidata:SPARQL_query_service/WDQS_backend_update/WDQS_backend_alternatives, JENA seems to take advantage concerning Overall technical assessments, but it's not the best choice concerning Server assessment by user criteria. Do you have hints about what the final decision will be? If not, when will it be decided and implemented? In Wikidata:SPARQL query service/WDQS backend update
https://wikidata.org/wiki/Wikidata:SPARQL_query_service/WDQS_backend_update you say that the deadline is 5 years since 2021.

Solution is based on a new search backend, but also in the implementation of a new code language.

2. Questions about search service accessibility and usage [Accessibility quality dimension]

Within the report “WDQS BACKEND ALTERNATIVES: THE PROCESS, DETAILS AND RESULT”, concerning the evaluation criteria "Query builder interface", you highlight two requirements for the new Wikidata Search Service after Blazegraph :

- *The interface provided at <https://query.wikidata.org/> includes a string based and templated builder*
 - *However, if a user interface is provided by the alternative offering, then this may be useful to replace or improve current capabilities*
1. Concerning the second item, do you envisage additional search interfaces to provide? Do they include customization devices?
 2. Is it possible to access metrics on the usage of the web query user interface? (i.e. Wikidata Query Service dashboard?)

3. Questions about Wikidata ontology

[https://commons.wikimedia.org/wiki/File:WikidataCon_2021_-_Overview_of_ontology_issues.pdf]

1. According to Vrandečić, Pintscher and Krötzsch (2023, p. 619): "While looking for funding, at least one major donor dropped out because the project proposal insisted that the ontology of Wikidata had to be community-controlled, and would be neither pre-defined by professional ontologists nor imported from existing ontologies." To what extent will this principle be maintained? Do you consider that the maintenance of this principle is the main cause of the semantic inconsistencies of the ontology?
2. In this same article, you say: “3.3. "Indeed, data uniformity and coherency has emerged as one of the big challenges that Wikidata has yet to address." What are the actions/projects you envisage concerning this point?
3. Also, in the article you say: "Another aspect of Wikidata that we think needs further development is how to more effectively share semantics – within Wikidata itself, with other Wikimedia projects, and with the world in general." What are the actions/projects you envisage concerning this point?
4. Also, in the article you say: “A variety of powerful knowledge representation languages could be used for this purpose, but we still lack both infrastructure and best practices to use them effectively in such complex applications.” Are you thinking in kind of taxonomy for this purpose? In this way, how do you assess labeling within the ontology (equivalence relationship) and subclass (taxonomic) relationships between classes?
Labeling: <https://wikidata-terminator.toolforge.org/#/>
Classification (taxonomy): It's an open question. We could work on it. They would appreciate a contribution in the solution.
5. One of the resources that have been created and adopted by Wikidata for recommending how to use its underlying ontology is “Property constraints”. Do you think it should be stronger (or not so flexible as nowadays) in order to improve accuracy, trustworthiness and consistency within the ontology?
6. Another of the aforementioned resources is Wikidata schemas, expressed with Shape Expression language (ShEx) https://wikidata.org/wiki/Wikidata:WikiProject_Schemas. Three of the *Important Entity Schemas* are:
 - E10 human (simple schema for humans)
 - E14 human (scheme for humans with basic properties)
 - E48 human (no description)

The first two include “property” P21 (gender) as a compulsory property, but not the third one. What is the degree of use of these entity schemas? Do you know other similar projects that can consolidate the gender perspective in Wikidata?

7. Within the report “WDQS BACKEND ALTERNATIVES: THE PROCESS, DETAILS AND RESULT”, concerning the evaluation criteria “Dataset evaluation by SHACL (SHape Constraint Language), ShEX (Shape EXpressions), or similar”, you highlight three requirements for the new Wikidata Search Service after Blazegraph :

- Although integrity checking is done prior to loading in the WDQS store, querying the store provides a secondary means to verify the data and find more complex errors
- If the backing store and query service can provide more advanced capabilities, then that is valuable
- Integrity can always be validated using SPARQL functions. However, newer languages (such as SHACL) may make this task easier.

Which kind of data integrity checking are you applying now? Are you thinking about designing "shapes graphs" for the SHACL-based validation? Further on: do you think it would be useful to set RDFS inferencing in the knowledge graph of Wikidata?

Most important actions to solve the mess in the ontology are:

- *Rules on properties (constraints)*
- *Survey among people about improving the ontology. Results will be published on July 2023. Source: https://commons.wikimedia.org/wiki/File:WikidataCon_2021_-_Overview_of_ontology_issues.pdf*
- *Entity schemas*

4. Questions about Wikidata and Wikipedia

Currently, there is a deep interrelation between WP and WD.

- Wikidata in Wikipedia: 74% of articles using Wikidata throw Infoboxes and Authority control; Petscan search system.... In addition, a Wikimedia project under current development asks for a deep relationship between Wikidata and Wikipedia: Abstract Wikipedia. DASHBOARD email Lydia PM Analytics questions
- Wikipedia i Wikidata: sitelinks; references...

1. How do you assess the current interrelation between Wikidata and Wikipedia?
2. In your opinion, what are further steps in these interrelations?
3. Do you think it's possible to get deeper in these interrelations so as to replace category schema by a unique ontology? What would be the Strengths, Weaknesses, Opportunities, and Threats of an initiative like this.

In the big picture, Wikidata ontology should replace category system of Wikipedia.

Concerning the gender gap, even in de.wikipedia.org there are conflicts with using or not the female gender within categories. For exemple: Koch vs. Köchin.

Wikidata ontology is not perfect, but it's a good alternative as a knowledge organization system for Wikipedia. And from the gender perspective is, definetely, better.

4. According to Vrandečić, Pintscher and Krötzsch (2023, p. 621) "The initial Wikidata proposal [...] was split in three phases: first sitelinks, second statements, third queries. The third phase, though, has not yet been realised. It was planned to allow the community to define queries, to store and visualise the results in Wikidata, and to include these results in Wikipedia. This would have served as a forcing function to increase the

uniformity of Wikidata's structure." Including results in Wikipedia can open the path to include also properties from Wikidata's ontologies?

Concerning queries. In the original proposal: about list articles in wikipedia, give me the context result from wikidata Bot <https://wikidata.org/wiki/Wikidata:Listeria>

For labeling, a key tool is <https://wikidata-terminator.toolforge.org/#/>

Annex B. Guió avaluació persones usuàries

Estic fent el doctorat al programa d'Informació i Comunicació de la UB i necessitaria que facis unes tasques a la Viquipèdia en català per ajudar-me a saber si l'organització dels continguts d'aquesta enciclopèdia s'adapta a les seves persones usuàries. En el teu cas, m'interessa que tu facis els exercicis que et demanaré perquè ets LECTOR/EDITOR.

Ens interessarà que **busquis una informació** que et demanarem dins l'enciclopèdia. I **no avaluarem la feina que fas tu**, en cap cas, el que avaluem és la Viquipèdia i com es dona accés als continguts. A més a més de buscar la informació, ens interessarà que mentre facis la tasca ens vagis **verbalitzant els teus pensaments**, com per exemple: necessitem que ens expliquis perquè fas clic en un lloc i no en un altre, que ens diguis perquè tires endavant o endarrera, que expressis els teus sentiments com ara si t'està costant o ho trobes fàcil, etc.

Necessitaria també que em signessis un consentiment informat per tal que puguem enregistrar l'àudio d'aquest exercici així com la pantalla (en el cas virtual) Consentiment informat: 2 còpies per persona (1 cal digitalitzar-la i desar-la a la carpeta test d'usuari un cop signada i 1 és per la persona voluntària) [CONSENTIMENT INFORMAT DE LA PERSONA PARTICIPANT.docx](#)

Per començar necessitaria que em comparteixis unes dades socio-demogràfiques:

1. Navegació

- **Vols fer un estudi de dones influents a l'educació durant el segle passat. Per això, buscaràs dones professores o mestres del segle XX.**

URL: <https://ca.wikipedia.org/wiki/Portada>

2. Cercadors

a. Cercador general:

Has fet servir mai el cercador de la Viquipèdia? Saps on és?

[Si no ho sap: El cercador general de la Viquipèdia es troba a dalt de tot de la pàgina, i aquest pugui servir per a fer una cerca o bé per trobar les paraules que et proposa el cercador a partir de la/les paraula/es que escriguis.]

- **Fem un estudi sota les dones a la ciència, tens la feina de trobar dones rellevants al món de la química però que siguin de Catalunya. Com ho localitzes a través del cercador de Viquipèdia?**

[Si no sap on és el cercador: URL: <https://ca.wikipedia.org/wiki/Portada>]

b. Cercador de gènere:

Saps que hi ha un cercador de gènere a Wikipedia. L'has fet servir mai? Saps on és?

El nom del cercador de gènere: què et sembla que vol voler dir?

La teva 1a impressió del cercador quina és? què et sembla que buscarà?

[Si no ho sap:

El cercador per gènere ha sigut el resultat de la deliberació de les categories per gènere a la Viquipèdia. Per a facilitar la cerca de dones i persones de altres gèneres han establert un cercador addicional en les pàgines de professions.

URL: <https://ca.wikipedia.org/wiki/Categoria:Pol%C3%ADtics>

- **Vols analitzar la presència de les dones al món de la política catalana del segle XX. I necessites fer una entrevista a una dona política catalana nascuda entre el segle XIX i XX, i que la seva activitat política s'hagi desenvolupat majoritàriament al segle XX per al teu estudi. Ho hauries de buscar a la Wikipedia.**

[Si no troba el cercador de gènere posem a la persona davant del cercador de gènere:
<https://ca.wikipedia.org/wiki/Categoria:Pol%C3%ADtics>]

c. Cerca interna

Saps que hi ha un cercador intern. L'has fet servir mai? Saps on és?

El nom del cercador intern: què et sembla que vol voler dir?

La teva 1a impressió del cercador quina és? què et sembla que buscarà?

[Si no ho sap:

El cercador de cerca interna es un més avançat, os l'usuari/a pot seleccionar més paraules i categories per trobar la cerca desitjada.

URL:

<https://ca.wikipedia.org/w/index.php?search=Professor&title=Especial:Cerca&profile=advanced&fulltext=1&ns0=>

- **Vols analitzar la presència de més gèneres a la Viquipèdia. Més específicament, persones no-binaries. En un primer moment, vols trobar un periodista no-binari. Com ho faries?**

d. Cercador “intersecció”

[Començar igual que amb els altres cercadors]

Què vol dir cerca interna? On és aquest cercador? Pots dir per a què serveix? Ho fas servir habitualment?

El nom del cercador: Intersecció què et sembla que vol voler dir?

Aquest és un cercador que disposa de Wikidata i més bases de dades per trobar cerques dins de la Viquipèdia.

URL:

https://petscan.wmflabs.org/?language=ca&project=wikipedia&categories=Pol%C3%ADtics&interface_language=ca

- **Has de fer un estudi de les dones a la ciència i la seva influència dins de l'educació. Per tant, vols buscar una professora de l'àmbit de STEM que sigui nascuda al llarg de l'any 1900 de qui puguis obtenir el màxim d'informació possible.**

3. Arbre de categories

Sabies que la Viquipèdia etiqueta els seus articles amb categories? Saps on són les categories als articles?

Saps que és l'arbre de categories? Fas servir habitualment les categories per localitzar informació?

a. Buscar l'arbre

- URL: <https://ca.wikipedia.org/wiki/Portada>
- Tasca:

Deus fer un estudi de les infermeres catalanes més importants del territori actualment, i vols fer-li una entrevista. Quina categoria de l'arbre hauries de fer servir per localitzar “infermeres”?

b. Buscar directament des de l'arbre

- URL: https://ca.wikipedia.org/wiki/Especial:Categories_en_arbre/Infermers
- Tasca:

Ara, a partir de l'arbre de categories, busques una infermera catalana per a fer-li una entrevista.

2. Índex i arbre

Saps a què és l'índex" a Viquipèdia? Has fet ús alguna vegada d'aquest? Ho coneixies abans que l'arbre de categories?

- a. Context: l'índex de categories es tracta de la pàgina que mostra les subcategories dins d'una concreta i la quantitat d'articles i categories dins de cadascuna.
- b. Dues tasques:

i. **Buscar per l'índex abans de l'arbre**

Tasca: Vols llegir més literatura feta per dones, però no coneixes moltes escriptores de rellevància a la llengua catalana a part de Mercè Rododera o Carme Laforet... les poques dones que has pogut estudiar a l'escola. Troba l'índex d'"escriptors" i busca una dona que sigui important a la llengua catalana entre els segles XX i XXI.

<https://ca.wikipedia.org/wiki/Categoria:Escriptors>

ii. **Buscar per l'arbre abans que utilitzar l'índex**

Tasca: Vols llegir més literatura feta per persones no-binàries però no coneixes a ningú ja que no es parla massa d'aquestes. Troba l'arbre de categories d'"escriptors" i busca una persona no-binària que sigui important a la llengua catalana entre els segles XX i XXI.

3. **Tests per a experts**

- a. **Generador de Consultes de Wikidata**

Coneixes aquests recurs? El fas servir? Si és que sí quan l'utilitzes, per quines necessitats d'informació?

Context: Aquest cercador es tracta d'un específic a Wikidata, una de les bases de dades amb les quals treballa Wikipedia. Aquest cercador també et permet trobar resultats més concrets de Viquipèdia. La base de dades funciona amb propietats dins de les quals es troben altres propietats d'instància que s'engloben dins d'altra, per exemple en 'Humà' pots trobar 'població' (P1082) o 'cultura' (P2596), però també pots trobar valors, com dins de 'nom' (P2561) pots tindre el valor de 'Yessica'. Cadascun dels valors té un identificador que generalment s'especifica amb 'Q' (ex.: QXXXXX).

Tasca: Has de fer un perfil d'un escriptor no-binari pel motiu del dia de la visibilitat LGBTI+. Per això, deus buscar una persona no-binària nascuda a Espanya per a trobar alguna de les seves obres més importants del moment. Decideixes fer-ho per Wikidata. Tracta de trobar des del generador de consultes de Wikidata. A partir d'ací, com ho fas?

URL: <https://query.wikidata.org/querybuilder/?uselang=ca>

- b. **Wikidata: Llista de propietats**

Context: les llistes de propietats mostren totes les categories per les quals classifica Wikidata el coneixement, tant per a la persona usuària com per al llenguatge de programació que fa ús la

base de dades i que emplen pàgines com la Viquipèdia. Per això, conèixer el llenguatge que utilitza Wikidata ajuda molt a la comprensió del funcionament de la Viquipèdia.

URL: https://wikidata.org/wiki/Wikidata:List_of_properties/ca

- i. Exploració de les propietats
 - **Tasca: T'han demanat un informe sota la possibilitat de gèneres que existeixen a la realitat, per fer ús de tots ells i ajudar a ampliar la visió de la teva empresa que fins ara va tindre una perspectiva binària. Has de trobar totes els que puguis a partir d'explorar en la llista de propietats de Wikidata.**
- ii. Exploració del cercador
 - **Tasca: Has de crear un llistat de dones escriptores. Des de Wikidata, has de buscar l'identificador d'escriptors buscant mitjançant el cercador.**
- iii. Combinació de les dues
 - **Tasca: Fes una consulta, relacionada amb gènere, que més interessant et sembli. Pots utilitzar tant el cercador com la llista de propietats.**

Annex C. Primera proposta d'avaluació de taxonomies

Criterios de evaluación				
Albuquerque (2017)	Kaplan et al. (2022)		Kless and Milton (2010) vocabularis controlats / Documentacio	
Marco conceptual	Generalidad	Concisión	Pureza conceptual	
			Exhaustividad conceptual	Exhaustividad del punto de vista
			Especificidad	
	Lucidez	Ausencia de redundancia conceptual		
		Claridad conceptual		
		Nivel de preordinación		

Performance/ Desempeño	Apropiación	Compleitud	Longitud/Cantidad de la representación del concepto	
			Exhaustividad terminológica	
		Solidez	Ausencia de redundancia terminológica	
			Corrección sintáctica	
Formato	Ortogonalidad	Corrección estructural	Corrección estructural	
			Pureza estructural	
Sistemas de ayuda		Navegabilidad		Longitud de la ruta de conceptos
				Alcance de la navegación
		Integridad de la documentación		
		Calidad de la documentación		
		Complejidad		
Consistencia				

Annex D. Primera proposta d'avaluació d'ontologies

Criterios de evaluación

Extrínsecos	Medida de las cualidades externas	Análisis de la calidad externa (estructura)		
	Aplicación	Diseño del contexto	Eficiencia	
			Accesibilidad	
			Disponibilidad	
			Recuperabilidad	
			Entendible/Claridad	
	Dominio	Diseño del contexto	Adaptabilidad	
			Precisión	
			Relevancia	
			Completo funcionamiento	
			Puntualidad/Conveniencia	Actualidad
			Volatilidad	
Credibilidad			Historia	
			Autoridad	
Intrínsecos	Cualidades del dominio intrínseco	Semántica del vocabulario	Concisión	
		Diseño de la arquitectura	Cobertura	

			Consistencia externa	
			Comprehensibilidad	
	Cualidades de la estructura intrínseca	Sintaxis	Cumplimiento normativo	
		Jerarquía	Complejidad	
		Diseño de la arquitectura	Consistencia interna	
			Modulación	