

Embodiment of an Agent by a Pepper Robot for Explaining Retrieval Results

Simon Schiff^{1,*}, Magnus Bender¹ and Ralf Möller¹

¹ University of Lübeck, Institute of Information Systems, Ratzeburger Allee 160, 23562 Lübeck, Germany

Abstract

Conceptually, an agent perceives its environment through sensors, builds a set of models, and then uses these models to select an appropriate action to fulfill its goals. As long as an agent is embodied by a robot, even humans that are not familiar with the concept of an agent, are more likely aware of the presence of an individual, independent of how the agent maps state sequences to actions, than if an agent is part of a web application. In the latter, agents are sometimes visualized as an animation, such as Clippy by Microsoft. Thus, depending on the context, it is often explicitly desired, that humans are aware of an individual, while they interact with a system. Our aim is to demonstrate the prototype of our information retrieval (IR) agent, running in the background of our information system (IS), implemented for humanities scholars. Instead of animating our IR agent, we embodied it by a Pepper robot for demonstration purposes only. Pepper is a humanoid robot especially designed for the interaction with humans, as he has among others a speech-to-text and text-to-speech module allowing for a verbal conversation between a human and the robot. We tested our approach with humans of which not everyone was familiar with the concept of an IR agent. During the interaction with our IS, Pepper explains, as the IR agent, his behavior. The embodiment of our IR agent, using Pepper, helps to understand the concept of an IR agent and that it is running in the background of our IS, without explaining that explicitly.

Keywords

Agent, Robot, Information Retrieval, Demonstration, Curated Datasets, Information System

1. Introduction

An agent in pursuit of a task perceives its environment through sensors, builds a set of models, and then uses these models to select an appropriate action to fulfill its goals [1]. It is perceived as being intelligent depending on which actions are selected, given the current state of its environment and its goals, regardless of which (artificial intelligence (AI)) methods are in use to map state sequences to actions. One of these goals could be for instance to satisfy the information need of a human. In this case, an IR agent, that has access to a large corpus of documents receives a query and its goal is to assign to each document in its corpus a score, given the query. Top n highest scored documents are returned to the human in descending

2nd Workshop on Humanities-Centred Artificial Intelligence (CHAI)


*Corresponding author.

✉ schiff@ifis.uni-luebeck.de (S. Schiff); bender@ifis.uni-luebeck.de (M. Bender); moeller@ifis.uni-luebeck.de (R. Möller)

🌐 <https://www.ifis.uni-luebeck.de/index.php> (S. Schiff); <https://www.ifis.uni-luebeck.de/index.php> (M. Bender); <https://www.ifis.uni-luebeck.de/index.php> (R. Möller)

🆔 0000-0002-1986-3119 (S. Schiff); 0000-0002-1854-225X (M. Bender); 0000-0002-1174-3323 (R. Möller)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

order. Assuming, the query and the documents are sequences of words, then functions such as TF.IDF, assigning a score to each query, document pair, have been shown to be effective in practice [2, 3]. However, such an IR agent, is not necessarily perceived as being intelligent.

For instance, given a corpus of fantasy novels and the query “bike repair shop”, an IR agent would return documents that are the most relevant ones with respect to the query, but not with respect to the information need of a human, that has a bike with a flat tire. The IR agent should approximate the true information need of the human from the query and the expectations the human has about the IR agent itself. The human expects to retrieve at least a document from the IR agent, containing a list of bike repair shops, of which the agent is unable to return. If the IR agent is able to identify the gap between the expectations of the human and its ability to satisfy its information need, then it can select an appropriate action not only given its goal to chose the most relevant documents, given the query. Additionally, it can act legible by explaining its behavior, if the gap is too large. That is a step towards to gain trust by the human and thus to be perceived as being intelligent. In this work, we implemented and evaluated our IR agent, as an extension to our IS, implemented as a web application [4].

The web application enables humanities scholars to upload Word documents for the creation of curated datasets. Uploaded Word documents are parsed, preprocessed, and depending on the context, split into several documents. For instance, a Word document containing hundreds of poems is split into a corpus of documents, where each document contains one poem. We have created for various types of documents, such as poems, viewer to view the contents of the documents at the web application. Additionally, links are created automatically, that help to jump between, for instance, words in poems and their corresponding entries in a dictionary.

Our IR agent, part of the web application, not only ranks uploaded documents, given a query, by relevance in descending order, it additionally returns for each document and score an explanation, in order to act legible. A web application, providing a search interface, is usually accessed by a human, using a tablet, smartphone, laptop, etc. does not expect to interact with an IR agent. Our aim is to evaluate our IR agent with real humans, who know that there is an IR agent, acting in the background. That can be solved by either explaining the concept of an IR agent, as we do in this paper or to embody our IR agent with a robot, having a text-to-speech module. The robot verbally explains its behavior on demand and humans are aware of that there is an IR agent running in the background, which changes their expectations and perceptions, while using our IS. We evaluate our approach, by using a Pepper robot [5]. As has been shown, the Pepper robot is a very effective tool to show to others what happens in the background of our IS system, without explaining explicitly the concept of an IR agent.

We introduce in Section 2 our IS that we extend with our IR agent we present in Section 3. In Section 4 we show how we use a Pepper robot to present our work to an audience, where some never heard of an IR agent before. Finally, we present related work in Section 6, conclude our results in Section 7, and give an outlook for future research directions.

2. Web Application

Humanities scholars work with specific tools and document formats across chronological and geographical borders to reach their goals. For instance, the goal is to produce a critical edition,

from a large collection of palm-leave manuscripts and editions, such as [6] created by Eva Wilden. A critical edition contains the trajectory a text made through various manuscript and print versions into the modern days. Producing a critical edition can take up to several years and often many humanities scholars are involved. Regardless of preferred document formats and tools in use, a finished critical edition is mostly published as a printed book or online as a PDF. We argue that this violates the FAIR (Findable, Accessible, Interoperable, Reuse) principles. Findable is often not a problem at all since published books have mostly associated metadata to be findable, by humans and machines. However, the contents of a critical edition are possibly not searchable and require a faceted IR system. Accessibility does not only account for of how the data is accessible, additionally it is important to make clear who is allowed to access what. For instance, not everyone is allowed to access some pictures of manuscripts in the printed books, but everything else. Only those who are allowed to see the images, are allowed to access the printed critical edition, as the images are inseparable from the rest of the book. A printed book or a PDF is made for humans to be readable and not to be interoperable with other programs except those that visualize or print the contents of a PDF. Finally, metadata should be well-described, such that other programs can reuse the associated data. A critical edition that does not violate the FAIR principles allows for faceted searches, automatic linking, access control, and the transformation of contents into various formats. However, humanities scholars prefer to use *what you see is what you get* (WYSIWYG) tools such as Microsoft Word, as they see always the current state of the book. Our web application allows humanities scholars to still work with their preferred tools, such as Microsoft Word, and document formats across chronological and geographical borders, and yet to produce data that does not violate the FAIR principles. Word documents can be uploaded at our web application, specific parts that are written in a specific controlled natural language, are parsed, split, and loaded into a database. The parser is automatically generated from an Antlr4 [7] grammar, allowing to be adapted easily to other types of documents. Viewer, part of the web application, are used in lectures for the visualization of the contents of the database. Additionally, one can merge specific parts of documents automatically on demand, which would take a humanities scholar weeks of work [4]. Among these features and those we would like to add in the future, we implemented an IR agent, we present in Section 3 and evaluate in Section 4, by embodying it by a Pepper robot.

3. Information Retrieval Agent

Word documents, containing hundreds of texts, such as poems, are treated each as corpora of texts, where each text is a document. Given a word and its context, part of a document, one could be interested in other documents, containing text snippets within the same context. We assume that the surrounding words of a word within a text make up the context and refer to the context to as *subjective content descriptions* (SCDs) [8]. Our IR agent assigns a score to all text snippets within all documents in the corpora, given a word and its context, as a query. Additionally, our IR agent adds an explanation for each score it assigns to the text snippets. Finally, all text snippets are returned to the human along with the associated document and an explanation in descending order sorted by score.

More formally, our IR agent has access to a set of documents D part of a corpus C . Each

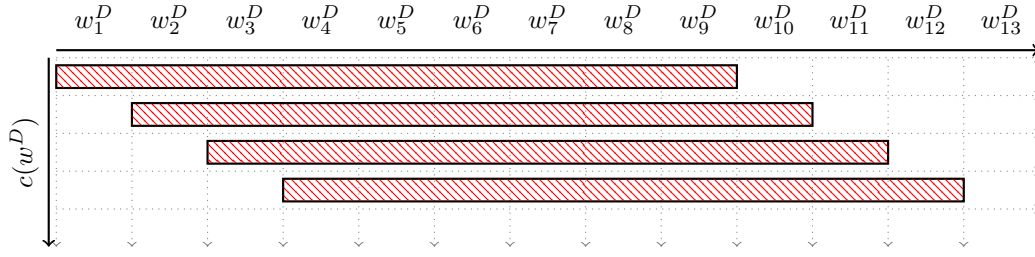


Figure 1: Window Function over a Sequence of Words $\langle w_1^D, \dots, w_{13}^D \rangle$ with $r = 4$

document D is a sequence of words $\langle w_1^D, \dots, w_n^D \rangle$ of length n . We assume that the surrounding words $\{w_j^D \mid i - r \leq j \leq i + r\}$ of a word w_j^D within a given radius r make up the context of the word w_j^D . As depicted in Figure 1, document D is a sequence of words $\langle w_1^D, \dots, w_{13}^D \rangle$. The context is highlighted in red cross lines, each for the words w_5^D , w_6^D , w_7^D , and w_8^D respectively. For instance, the words that make up the context for the word w_7^D are $\{w_3^D, \dots, w_{11}^D\}$, as depicted in the third row in Figure 1. In Algorithm 1, we show how to initially compute for each word in every document in the corpus, the words, that make up the context. The result is a mapping c that maps all words $w_j^{D_i}$ to a set of words, that make up the context: $c : w_j^{D_i} \rightarrow \{w_{j-r}^{D_i}, \dots, w_{j+r}^{D_i}\}$, with D_i being the current document, j being the position of word $w_j^{D_i}$ in document D_i , and r the radius. All sets of words in one document, have possibly

Algorithm 1 Compute Windows

```

1: procedure CONTEXTWINDOWS( $C, r$ ) ▷ Corpus  $C$  and radius  $r$ 
2:    $c : w_j^{D_i} \rightarrow \{w_{j-r}^{D_i}, \dots, w_{j+r}^{D_i}\}$ 
3:   for all  $D_i \in C$  do
4:     REMOVESTOPWORDS( $D_i$ ) ▷ Remove stop words from document  $D_i$ 
5:     for  $j \leftarrow r$  to  $|D_i| - r$  do ▷ Length  $|D_i|$  of document  $D_i$ 
6:        $c(w_j^{D_i}) \leftarrow \{\}$ 
7:       for  $k \leftarrow j - r$  to  $j + r$  do
8:          $c(w_j^{D_i}) \leftarrow c(w_j^{D_i}) \cup \{w_k^{D_i}\}$ 
9:   return  $c$  ▷ Return mapping  $c$ 

```

similar sets in other documents. We measure the similarity of two sets by the size of their intersection (i.e. the number of words they have in common). If it is above a given threshold t , then we assume that both contexts are similar up to an extend.

Algorithm 2 returns a mapping r , mapping words in all documents to text snippets in other documents, from the same context, if the similarity is above a given threshold t . A human, that is interested in text snippets from a similar context, sends a word as a query to our IR agent. Our IR agent returns all documents of similar context, given a word as a query, that contain text snippets returned by mapping r , with respect to the similarity of the text snippets in descending

Algorithm 2 Compute Results

```
1: procedure COMPUTERESULTS( $c, t$ ) ▷ Contexts  $c$  and threshold  $t$ 
2:    $r : w_j^{D_i} \rightarrow \{c(w_1^{D_1}), \dots, c(w_l^{D_k})\}$ 
3:   for all  $w_j^{D_i} \in c$  do
4:     for all  $w_l^{D_k} \in c$  do
5:        $i \leftarrow |c(w_j^{D_i}) \cap c(w_l^{D_k})|$ 
6:       if  $D_i \neq D_k$  and  $i \geq t$  then
7:          $r(w_j^{D_i}) \leftarrow r(w_j^{D_i}) \cup c(w_l^{D_k})$ 
8:   return  $r$  ▷ Return mapping  $r$ 
```

order.

Given the j -th word $w_j^{D_i}$ in document D_i , the context of the word $c(w_j^{D_i})$, a text snippet $c(w_l^{D_k}) \in r(w_j^{D_i})$, with $(w_l^{D_k}) = \{w_{l-r}^{D_k}, \dots, w_{l+r}^{D_k}\}$, from another document $D_k \neq D_i$, and radius r , our IR agent has to generate an explanation, of why it has returned the document D_k , among others, as a result for the query $w_j^{D_i}$. It generates for each document D_k in the result set, an explanation, by returning an excerpt from each document, that contains the words $c(w_l^{D_k}) \in r(w_j^{D_i})$. Each excerpt is a sequence of words containing $c(w_l^{D_k})$, of which $c(w_j^{D_i}) \cap c(w_l^{D_k})$ are emphasized. The human can decide to send another query to the IR agent, and to change radius r or threshold t . Even if results do not satisfy the information need of the human, the IR agent acts legible from the perspective of the human. Changing r and t each to a value that leads possibly to more sophisticated results, is possible by the human, as the IR agent explains of how it computes a result.

4. Evaluation

At the open day of the Centre for the Study of Manuscript Cultures (CSMC)¹ we presented our web application to an audience, at where not everyone is familiar with the concept of an IR agent. Instead of explaining the concept of an IR agent, we embodied the IR agent by the Pepper robot [5]. The experimental setup is depicted in Figure 2. The presenter sits in front of a table with a laptop, hosting the web application as well as running a web browser for accessing the web application. A 75 inch large screen is behind the presenter, in a height such that the whole screen is visible for all visitors in front of the table, mirroring the screen of the laptop. Pepper stands on the left hand site of the table, near enough for the visitors to see the contents of the tablet on its chest and to hear what he says. Our web application is controlled by the presenter.

In addition to various sensors and actuators, Pepper has a machine inside his head and an Android tablet attached to his chest. The machine in his head is equipped with a quad-core Intel Atom E3845 processor up to 1.91 GHz, 4 GB RAM, and a flash memory of 32 GB. An Android tablet, connected via an internal network with the machine in Peppers head, has a 10.1 inch

¹<https://www.csmc.uni-hamburg.de/openday-en.html>

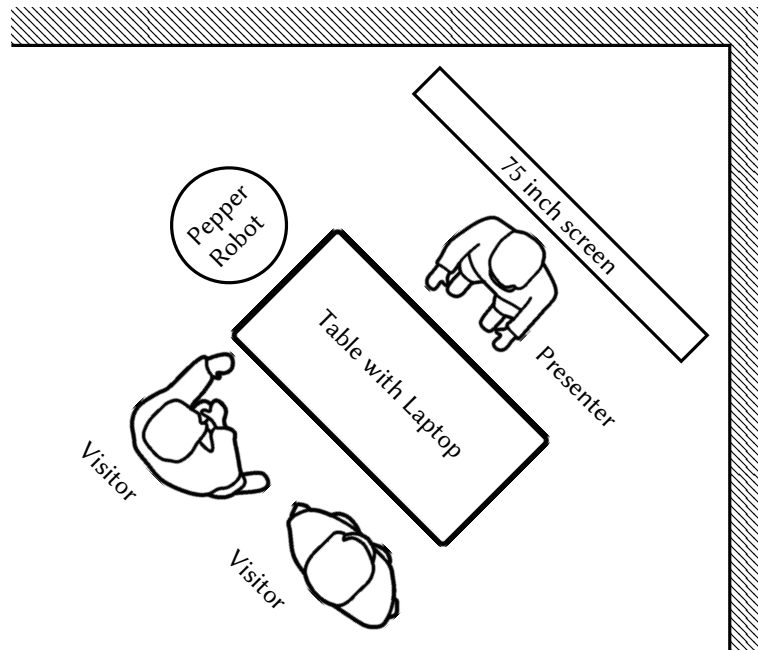


Figure 2: Experimental Setup of our Evaluation

display, a TCC8925 processor with a single ARMv7 A5 core up to 833 MHz, and 1 GB RAM [5]. It currently runs Android 6.0 “Marshmallow”, allowing to install Android apps from the official Google Play store and to deploy self developed Android apps. Due to the hardware limitations of the tablet, even the Android interface itself is sometimes jerky, therefore the graphical design of Android apps is limited up to an extend.

Pepper is equipped with a text-to-speech module, that can be accessed over an application programming interface (API), when one develops an Android app, that runs on the tablet of Pepper. The tablet, then sends texts over the internal network to the machine inside Peppers head, that is responsible, among others, for translating text into speech, that then the human can hear over Peppers speakers. As depicted in Figure 3, the laptop, running the web application, is connected with Pepper over a network. We developed an Android app, that opens a WebSocket in the background, for receiving texts from a JavaScript interface, accessible using our web application. Texts are then forwarded over the API to the text-to-speech module inside Peppers head. As we added a web view to our Android app, our web application can be used both on the laptop and directly on the tablet of the Pepper robot. We have created a video of pepper for demonstration purposes.²

Approximately 20 visitors, from the humanities, chemistries, biologies, and computer sciences, have visited our stand at the open day of the CSMC. Only the computer scientists have heard of the concept of an IR agent at beforehand. The presenter uses the web application on the laptop to first upload a Word document. Pepper then explains how he processes the uploaded document, as if he is the IR agent in the background, as described in Section 2. After the document is

²https://www.fdr.uni-hamburg.de/record/10769/files/KI2022_CHAI-presentation4.zip

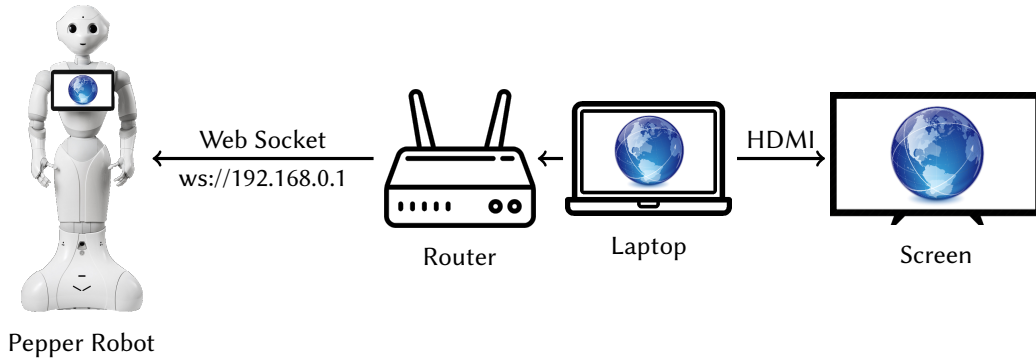


Figure 3: Architecture of the Demonstration

processed, its title is visible at the web application. It possibly consists of several texts, that are each treated as a document and loaded into a database. All documents in the database can be listed and its contents can be viewed with a viewer at our web application. Words $w_j^{D_i}$ with $r(w_j^{D_i}) \neq \emptyset$ are highlighted at the web application as to be clickable, while the others $w_l^{D_k}$ with $r(w_l^{D_k}) = \emptyset$ are not. The visitors decide on which word the presenter should click. Finally, Pepper as the IR agent, explains how it computes the results, as described in Section 3. As far as we can tell from feedback and questions in return to our presentation, all of the visitors were able to understand, of how our IR agent computes the results, given a query, that our IR agent is running in the background, and that the results are relevant. It was not necessary to explain all the technical details, as we do in Section 2 and Section 3.

5. Human Aware IR Agent

Visitors are aware of an IR agent, running in the background of our IS, implemented as a web application, as we embodied our IR agent by a Pepper robot. The Pepper robot explains as the IR agent, of how it processes documents and returns them sorted descending by a score it assigns to each of them, given a query. As we propose in [9], our IR agent can greatly improve its performance, if it would be aware of the human, such that they then can collaboratively seek for information. We refer to such an IR agent to as a human-aware IR agent, at where the human and the IR agent are modeled with their mental models \mathcal{M}^H and $\widetilde{\mathcal{M}}^A$ respectively, as depicted in Figure 4. On the left hand side, the IR agent approximates the information need of the human \mathcal{M}^H as $\widetilde{\mathcal{M}}_a^H$. However, the IR agent has its own mental-model $\widetilde{\mathcal{M}}^A$, containing the information need of the human, from the perspective of the IR agent. This is comparable to a customer explaining to an IT-specialist what requirements an application to be developed has to meet. The IT specialist has years of experience, identical to our IR agent that is able to go through all documents in a corpus it has access to, and knows that the program has to meet more than the customers requirements to work properly.

A human, sending a query to our IR agent, is aware of our IR agents mental-model $\widetilde{\mathcal{M}}^A$ and the IR agent itself is aware of that, as depicted on the right hand side in Figure 4. The human

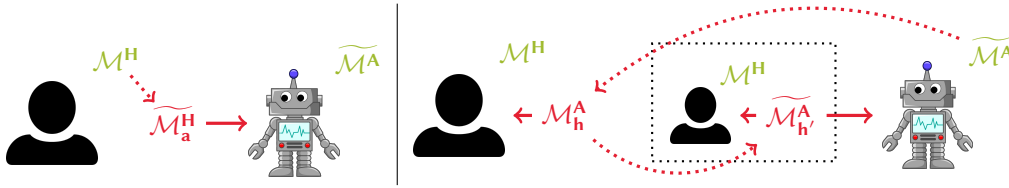


Figure 4: Mental models of the human \mathcal{M}^H and the IR agent $\widetilde{\mathcal{M}}^A$ [10]

approximates $\widetilde{\mathcal{M}}^A$ as \mathcal{M}_h^A and the IR agent approximates \mathcal{M}_h^A as $\widetilde{\mathcal{M}}_h^A$. If the gap between \mathcal{M}_h^A and $\widetilde{\mathcal{M}}_h^A$ is too large, then the IR agent's behavior is not explicable and it should explain its behavior. As in the example before, a human and an IT specialist aim to find all requirements a program has to meet. The human expects that the IT specialist has experience in developing an application and possibly expects suggestions for improvements. If the IT specialist notes, that the human does not understand his suggestions, then the specialist should explain them. The human is more likely aware of $\widetilde{\mathcal{M}}^A$ if the IR agent is embodied by a robot or animated, as we have shown in Section 4.

6. Related Work

The animation of an agent is often done to make humans aware of that an actual agent is running in the background, which can improve the collaboration between humans and agents and to make agents more life-alike [11, 12]. Even humans are more likely aware of the copresence of other humans, if they are animated as an avatar [13]. However, as has been shown in the past, the animation of an agent is not sufficient at all, as it has turned out with Clippy [14]. Among other things, Clippy often interrupts a person to provide assistance even though no help is needed and even if needed, the goals of humans are often wrongly anticipated. As Kambhampati et al. note in [10], this problem has not yet been solved in the field of robotics, where agents are embodied by a robot, but crucial for the collaboration between a human and a robot. Li et al. note in a survey that humans perceive agents more positively, when they are embodied by a robot that is physically on site rather than virtually present or animated [15]. Thellman et al. in [16] add that there might be no difference, with respect to of its social presence, but note that their study is domain-specific and short. Our contribution is to first make humans aware of our IR agent, running in the background of our web application and then, as a future work, to make our IR agent aware of the humans interacting with it.

7. Conclusion and Future Work

As has been shown at our demonstration, we do not need to explicitly explain the concept of our IR agent, if it is embodied by a humanoid robot, such as Pepper. Currently, we use the API of Pepper, such that it speaks out what the web application sends to it. The API provides more than that and we aim to extend our IR agent demonstration, such that visitors can interact with

it using the speech-to-text and text-to-speech modules inside the machine of Peppers head. That allows for perceiving our IR agent even more as an individual, that aims to collaboratively seek together with humans for information and thereby to satisfy their information needs.

As mentioned in Section 5, an IR agent can greatly improve its performance if it is human-aware. We will further develop our IR agent [9], such that it is human-aware and then can be embedded in the Pepper robot for demonstration purposes.

References

- [1] S. Russel, P. Norvig, *Artificial Intelligence: A Modern Approach*, 2021.
- [2] K. S. Jones, A statistical interpretation of term specificity and its application in retrieval, *Journal of documentation* 28 (2004).
- [3] J. Beel, B. Gipp, S. Langer, C. Breiting, Paper recommender systems: a literature survey, *International Journal on Digital Libraries* 17 (2016) 305–338.
- [4] S. Schiff, S. Melzer, E. Wilden, R. Möller, TEI-Based Interactive Critical Editions, in: *International Workshop on Document Analysis Systems*, Springer, 2022, pp. 230–244.
- [5] A. K. Pandey, R. Gelin, A Mass-Produced Sociable Humanoid Robot, *IEEE Robotics & Automation Magazine* 25 (2018) 40–48.
- [6] E. Wilden, A Critical Edition and an Annotated Translation of the Akanānūru: Part 1, Kalirriyānainirai. Old commentary on Kalirriyānainirai KV - 90, word index of Akanānūru KV - 120, *École Française d’Extrême-Orient*, 2018.
- [7] T. Parr, *The Definitive ANTLR 4 Reference*, Pragmatic Bookshelf, 2013.
- [8] F. Kuhr, T. Braun, M. Bender, R. Möller, To Extend or not to Extend? Context-Specific Corpus Enrichment, in: *Australasian Joint Conference on Artificial Intelligence*, Springer, 2019, pp. 357–368.
- [9] S. Schiff, R. Möller, On Human-Aware Information Seeking, in: *CHAI@KI*, 2021, pp. 31–39.
- [10] S. Kambhampati, Challenges of Human-Aware AI Systems, *CoRR abs/1910.07089* (2019). URL: <http://arxiv.org/abs/1910.07089>.
- [11] T. Holz, M. Dragone, G. M. O’Hare, Where Robots and Virtual Agents Meet, *International Journal of Social Robotics* 1 (2009) 83–93.
- [12] M. Thiebaux, S. Marsella, A. N. Marshall, M. Kallmann, Smartbody: Behavior Realization for Embodied Conversational Agents, in: *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, 2008, pp. 151–158.
- [13] M. Gerhard, D. Moore, D. Hobbs, Embodiment and copresence in collaborative interfaces, *International Journal of Human-Computer Studies* 61 (2004) 453–480.
- [14] N. Baym, L. Shifman, C. Persaud, K. Wagman, Intelligent Failures: Clippy Memes and the Limits of Digital Assistants, *AoIR Selected Papers of Internet Research* (2019).
- [15] J. Li, The Benefit of Being Physically Present: A Survey of Experimental Works Comparing Copresent Robots, Telepresent Robots and Virtual Agents, *International Journal of Human-Computer Studies* 77 (2015) 23–37.
- [16] S. Thellman, A. Silvervarg, A. Gulz, T. Ziemke, Physical vs. Virtual Agent Embodiment and Effects on Social Interaction, in: *International conference on intelligent virtual agents*, Springer, 2016, pp. 412–415.