Designing a Revenue Mechanism for Federated Search Engines

Laura Bonetti
DEI, Politecnico di Milano
piazza Leonardo da Vinci
Milano, Italy
laura.bonetti@mail.polimi.it

Sofia Ceppi
DEI, Politecnico di Milano
piazza Leonardo da Vinci
Milano, Italy
ceppi@elet.polimi.it

Nicola Gatti
DEI, Politecnico di Milano
piazza Leonardo da Vinci
Milano, Italy
ngatti@elet.polimi.it

ABSTRACT

Federated search engines constitute a new class of search computing paradigms whereby a multi–domain query is decomposed in a number of single–domain queries, each one addressed by a domain–specific content service provider. The paradigm provides a number of advantages, in particular the possibility to discover more pertinent information by scouring the *deep Web* and to find automatically correlations among the service providers' results. In our work, we focus on the design of a revenue mechanism for such paradigm. This task, although being of extraordinary importance, has not received enough attention in the literature so far. In particular, in this paper we describe a revenue mechanism in terms of *business model*, specifying who pays and when, and *microeconomic model*, specifying how the optimal payment can be computed. Futhermore, we discuss its properties.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics; H.3.5 [**Online Information Services**]: Commercial Services

General Terms

Design, Economics

Keywords

Online Advertising, Federated Search Engines, Revenue Mechanisms, Microeconomics

1. INTRODUCTION

The main general–purpose search engines crawl the Web and index Web pages, finding the best pages for each specific list of keywords with good precision. However, the so–called *deep Web* contains information that is largely more valuable than the one that a current general–purpose search engine can discover. The development of new searching paradigms able to address more complex searches than those addressed to the current search engines and to discover deeper information is currently one of the most interesting challenges in the search computing field. Currently, the emerging paradigm is based on the combination of a *multi–domain* query approach with the integration of heterogeneous data sources capable to scour the deep Web. This has resulted in a new generation of search paradigms, called *federated search engines* (FSEs), that integrate search results from heterogeneous domain–specific content service providers [20].

Consider for instance a user that searches for a restaurant close to a given hotel in a given city. General–purpose search engines allow the user to make exclusively single–domain queries on restaurants or hotels. Then, the user must manually scour the results of the query on restaurant and of the query on hotel to find a restaurant and an hotel close together. Instead, allowing a user to make multi–domain queries, the user needs to make only a single query in which she specifies all the fields of the search. The results she obtains contain both hotels and restaurants that are displayed in a way that helps the user to select the closest ones.

While the scientific community has been working on the development of FSE enabling technologies for some years [4], the problem of designing a suitable *revenue mechanism* has not yet received enough attention. This issue is of paramount importance for the success of the FSE paradigm. In the present paper, we focus on this problem. The starting point of our work is the analysis of the revenue mechanisms currently adopted in the search applications. A revenue mechanism is composed by a *business model*, that specifies who pays and when, and by a *microeconomic model*, that specifies how the optimal payment can be computed. Both models play a crucial role in the design of a successful revenue mechanism and depend on each other, e.g., insights from the microeconomics can be used to design the business model that allows one to maximize the revenue. Since many different revenue mechanisms are currently used, we propose a classification, emphasizing actors and payment schemes.

Then, we focus on the FSE scenarios. They involve many actors and require the design of a new revenue mechanism. This is because existing revenue mechanisms are not able to integrate different advertising services and to provide the actors involved in the FSE's activity with appropriate monetary incentives. We propose an *ad–hoc* revenue mechanism, stating initially its requirements and subsequently designing a business model and a microeconomic model that satisfy the requirements. Both models extend the existing ones. We also report an example of application and some very preliminary experimental results with Yahoo! dataset.

The paper is structured as follows. In the next section, we classify the existing revenue mechanisms for search applications. In Section 3, we present our business model for FSEs, while, in Section 4, we describe the microeconomic model. Section 5 concludes the paper.

2. REVENUE MECHANISMS FOR SEARCH APPLICATIONS

We provide a classification of the currently adopted revenue mechanisms in the attempt to provide a classification in terms of business models and microeconomic models.

2.1 Business Models

A business model can be characterized in terms of *actors* and *payment schemes.*

Actors can be classified as follows. *User*: her aim is to search for contents/advertisement. *Advertiser*: its aim is to appear in sponsored links or banners displayed to targeted users. *Content service provider*: its aim is to produce a list of contents (organic search), given an input provided by the user. The input can be a list of keywords or a more structured query. Contents are usually extracted from databases. As in the case of Zillow [26], that provides contents about houses to rent and sell, and of Expedia [10], that gives links (e.g., flights and rooms) related to tourism. Contents can be displayed by the provider itself or by another actor. *Advertising service provider*: its aim is to produce a list of ads (sponsored links) that target at best the user given an input. The input can be a list of keywords or a list of contents (e.g., web pages) and it is used to target the ads, as in the case of AdSence [1] (i.e., Google's advertising service provider [12]). Ads can be displayed by another actor that uses the service. *Integrated content and advertising service provider*: its aim is to produce a list of contents and a list of ads. The list of ads is chosen to target at best the user. The input is a list of keywords, e.g., in the case of Google that displays organic links and sponsored links using AdWords [2]. *Content integrator*: its aim is to integrate the organic search results of different content service providers. Very simple examples are meta search engines, e.g., Ecosia [7], that integrates content from Yahoo! [25] and Bing [3], and Ixquick [14] that uses links from Google, Yahoo!, Exalead [9], Wikipedia [24] and a lot of others.

The payment scheme defines when (usually in terms of events) an actor must pay another actor. The payment schemes usually adopted in search applications are the following. *Pay–per–query*: a payment is required to have access to data. *Pay–per–impression*: a payment is required to have a link displayed. *Pay–per–click*: a payment is required when a given displayed link is clicked. *Pay–per–conversion*: a payment is required when a conversion (e.g., a transaction) is accomplished.

Now we describe the basic business models in terms of actors and payment scheme (search applications can use more than one basic business model, as we discus below).

Business model 1: captures the situation in which a user needs to access data of a content service provider. The actors are a user and a content service provider, while the payment scheme is pay–per–query. Exactly, the user pays an amount of money per access to data to the content service provider. In Zillow, a real estate search engine, a user, who wants to buy a house, should enter the city as an input,

for example San Francisco. Zillow returns a list of houses with related information. If the user wants to obtain more details on houses and/or wants to make an offer for a purchase, she must register and pay Zillow. Some variations (e.g., *pay–per–thousands*) include the case in which a user pays a given amount of money for a given amount of accesses. This business model can be applied also when a content integrator acts in the place of the user.

Business model 2: captures the situation in which a user makes a search on an integrated content and advertising service provider. The actors are a user, an integrated content and advertising service provider, and some advertisers, while the payment scheme is pay–per–click. Exactly, the integrated content and advertising service provider displays contents and advertisement and, when the user clicks on an ad, the corresponding advertiser pays a given amount of money to the service provider. Google and its advertising service provider AdWords have this business model. If a user enters a keyword, for example flights to Milan, the search engine returns a list of organic links, as airlines sites, and some ads, for example content service providers for hotels (e.g., Booking.com). When a user clicks on one of the ads the related advertiser pays Google.

Business model 3: captures the same situation of the previous business model except that the content service provider uses an external advertising service provider. The actors are a user, a content service provider, an advertising service provider, and some advertisers, while the payment scheme is pay–per–click. Exactly, the content service provider displays contents it generates and the advertisement generated by the advertising service provider. When the user clicks a sponsored link, the corresponding advertiser pays a given amount of money to the advertising service provider that, in its turn, pays a ratio to the content service provider. Vivisimo [23] has this business model. If a user enters a keyword, for example hotels in San Francisco, the search engine returns a list of organic links, as hotel websites, and some ads, for example booking search engines as Edreams [8] and Trivago [22]. These links come from Google AdSense. When a user clicks on one of these ads the related advertiser pays AdSence for the received click and Vivisimo receives a percentage of this payment.

Business model 4: captures the situation in which a user searches for information to make a conversion and no advertisement is involved. The actors are a user and a content service provider, while the payment scheme is pay–per–conversion. Exactly, when the user has the conversion (e.g., a purchase), she pays a given amount of money to the content service provider. This is the business model used by Expedia [10]. We assume that the user wants to book a hotel. He introduces the city, such as San Francisco, and chooses a date, for example, from August 6–th to August 11–th. Expedia returns free rooms. If the user decides to book a room the site charges a reservation fee.

Business model 5: captures the situation in which a user searches for information to make a conversion and only advertisement is involved. The actors are a user, an advertising service provider, and some advertisers, while the payment scheme is pay–per–conversion. Exactly, when the user has the conversion (e.g., a purchase), the advertiser pays a given amount of money to the advertising service provider. Jellyfish uses this business model. We assume that the user wants to buy a pair of shoes and she inserts 'sneakers' as

keyword. Jellyfish returns a number of products consistent with the search. If the user decides to buy the product, the seller pays Jellyfish.

Actors may use more than one of the above business models. An example is Tripadvisor [21] that merges business model 3 and business model 5. A user can search for hotels, flights, restaurants and activities. We assume that the user wants to find a room in San Francisco from August 6–th to August 11–th. Tripadvisor returns hotels with a free room in the specified period. The used business model is business model 5, i.e., if the user books a room, Tripadvisor receives a payment from the advertiser. Moreover, for each search, ads are also displayed. In this case, Tripadvisor also adopts business model 3.

2.2 Microeconomic Models

Essentially two general microeconomic models apply to the previous business models: *pricing models* and *auction models*. Both models are based on mathematical optimization theory and, specifically, on game theory [11]. Pricing models define the best price at which a service can be sold and they are usually based on oligopolies [11]. These models capture static situations (e.g., in business model 1 the price required by the content service provider is fixed) and the associated optimization problem is solved offline. Auction models define how some resources are allocated and how much the player who receives a resource must pay. These models capture dynamic situations (e.g., in business model 3 the ads displayed depend on the specific search of the user and the budget of the advertisers) and the associated optimization problem is solved online. The fact that auction models must be solved online makes the study of these models more interesting. Due to this we limit our survey to them.

In online search applications, auctions are used by advertising service providers also for deciding which ads to display. Different kinds of online advertisement can be used [16]. The main ones are the following. *Banner ads or display ads*: an ad is a long thin strip of information that may be either static or may include a hyperlink to the advertiser's web page; *rich media*: the banner ad is enriched by streaming video, audio, and interactivity that can allow users to view and to interact with products and services; *sponsored search*: advertisers pay to be listed and linked when a specific word or phrase is searched by a user.

We describe the general model of an auction for advertising. An auction $\mathcal{M}$ is defined as a tuple

$$\mathcal{M} = \langle A, X, \langle \Theta_a \rangle_{a \in A}, f, \langle p_a \rangle_{a \in A} \rangle$$

where $A = a_1, \ldots, a_m$ is the set of advertisers; X is the set of possible allocations of ads on the web page, where an allocation x defines which ads are displayed by the advertising service provider and in which order; Θ_a is the set of possible bids that advertiser a can submit to the advertising service provider, usually $\Theta_a = \mathbb{R}^+$; $f : (\Theta_{a_1} \times \cdots \times \Theta_{a_m}) \to X$ is the social choice function, i.e., given the bids submitted by all the advertisers, the social choice function determines the allocation; p_a is the payment rule defining the payment of advertiser a. Essentially, an auction puts in competition all the advertisers that are interested in being displayed for the same query. In the case of Google and AdWords, advertisers register their ads for a given set of keywords and, every time a user searches for such keywords, an auction among the registered advertisers is carried out.
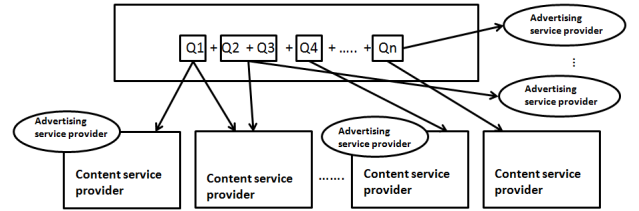

Figure 1: Reference scenario for a FSE's instance (for simplicity, users and advertisers are omitted).

The most studied auctions in literature are *efficient* auctions, defined as auctions in which f returns the allocation that maximize the sum of the bids of displayed ads. Different rules p_a can be found. The most known, when there is only one resource to allocate (e.g., a single banner or a single ad slot), are the *first price* auction, in which each advertiser pays an amount of money equal to its bid, and the *second price* auction, in which advertiser pays an amount of money equal to the second highest bid. Generalizations of them, for the case in which the resources to allocate are more than one, are the *generalized first price* (GFP) and the *generalized second price* (GSP). The GFP is the first auction mechanism adopted by Overture in 1997 [16]. Nowadays this mechanism is no longer used, all the main general–purpose search engines (e.g., Google) adopt the GSP mechanism [16].

Another auction model for scenarios with multiple resources is the *Vickrey–Clarke–Groves* (VCG) auction. Its payment rule is similar to the GSP's one, but it presents better microeconomic properties.

When the pay–per–click payment scheme is used (this is the most frequent case), refinements of these auction models are required. More precisely, each ad is associated with a *click probability* (denoted by q_a and estimated by the advertising service provider) that can depend on the user's search (e.g., the searched keywords), the specific ad, the position in which the ad is displayed, and the other displayed ads. Usually, to have a better targeted advertisement, a user model in which the user scans the ads from the top to the bottom, is considered.

3. A BUSINESS MODEL FOR FEDERATED SEARCH ENGINES

The federated search engine paradigm is based on the possibility for users to make multi–domain queries that are decomposed in multiple single–domain queries that are addressed to a domain–specific content service provider. An FSE is essentially a more sophisticated content integrator that exploits (and integrates) existing actors (those described in Section 2.1). In our mind, different instances of FSEs are possible. The organization of each one strictly depends on the specific application. Our attempt is the design of a reference scenario, specifying actors and their interaction, that can be applied to every instance of FSE. On the basis of this scenario, we design our business model.

Reference scenario. The scenario, depicted in Fig. 1, is characterized by the following actors: a user, the FSE, m content service providers, k advertising service providers, h integrated content and advertising service providers, and some advertisers.

In this scenario, a user inserts into the FSE a multi–domain query composed by n keywords $Q_1 + \cdots + Q_n$. The

FSE decomposes the multi–domain query into single–domain queries and addresses each of them to the pertaining content service providers and/or advertising service providers and/or integrated content and advertising service providers. For example, keyword Q_1 composes the single–domain query of the first integrated content and advertising service provider; it is also communicated to the first content service provider whose single–domain query is composted by keyword Q_1, Q_2, and Q_3. Each service provider communicates to the FSE its results. If it is a content service provider, it communicates its organic search results and information to combine them in an effective way; if it is an advertising content service provider, it communicates its ads and information to merge them in an efficient way; if it is an integrated content and advertising service provider, it communicates data and information about both organic search results and ads. When the FSE receives the results from all the service providers, it integrates the organic search results in an interrelated way and merges the ads lists. Then it displays them to the user.

Requirements. The above scenario requires an intricate business model due to the large number of actors. We define three main requirements for the business model.

Heterogeneity: the business model must integrate different (heterogeneous) basic business models (typically, those described in Section 2.1), adopting, for each actor, the most appropriate one.

Redistribution: the business model must assure (by means of revenue redistribution) that every actor, that takes part to the FSE's activity, receives an appropriate monetary incentive.

Flexibility: the business model must be (flexibly) tailored according to the specific online search application.

Model design. In order to satisfy the above requirements, we design a business model that combines some basic business models described in Section 2.1 with a new business model. The heterogeneity is addressed by allowing the co-existence of multiple basic business models. More precisely, we expect that the interaction between FSE and the other actors will be defined as follows:

- when FSE uses a content service provider to have access to data, the payment scheme is pay–per–query and therefore business model 1 is used;
- when FSE uses an advertising service provider, the payment scheme is pay–per–click or pay–per–conversion and therefore business models 3 or 5 is used;
- when FSE uses an integrated content and advertising service provider, the payment scheme is pay–per–click or pay–per–conversion and therefore business models 3 or 5 is used.

The above combination of business models does not address the redistribution requirement, not assuring all the actors to receive a monetary incentive to participate to the FSE's activity. The crucial issue concerns the merge of multiple integrated content and advertising service providers. The above business models prescribe that the advertiser whose ad has been clicked pays the service provider to which it is registered and the service provider gives a part of the revenue to the FSE. According to business model 3, all the ads generated by the advertising service provider must be displayed, but this is not possible when many service providers are used due to space limitation on the web page. In addition, the advertisement of one provider could result much more pertinent to the queries, affecting negatively the click probabilities of the advertisement of the other providers. As a result, this business model does not assure that all the service providers receive a monetary incentive to take part to the integration, therefore a provider could reject it. Furthermore, we should grant that an integrated content and advertising service provider gains more than content service providers, otherwise it would change the business model, providing only contents. What we need is to guarantee that:

- when a user clicks on an ad, the payment must be redistributed over all the involved actors assuring that each actor receive the appropriate monetary incentive to take part to the FSE's activity.

The appropriateness of the monetary incentives is defined by means of constraints (on the basis of commercial contracts) that are considered in the definition of the microeconomic model, e.g., assuring a given (minimal) revenue to each provider. Essentially, we assume that the revenue of each actor involved in the FSE's activity will be composed of a minimum fixed amount of money from each query plus a part that depends on the specific displayed ads.

The flexibility requirement can be addressed by allowing the change of the actors and their interactions. This requirement can be directly addressed by designing a flexible microeconomic mechanism.

Example. We report now an example, depicted in Fig. 2, of future application of FSE. The user Valentina is interested in planning her summer holiday in San Francisco. She wants to book flights from/to Milano to/from San Francisco, and she wants to find an hotel and few restaurants in San Francisco. Moreover, Valentina has decided to do bungee jumping as extreme experience during the holiday, and she wants to find a place where doing it. To obtain all the information she needs to plan her holiday, Valentina makes a multi–domain query composed by the following keywords:

- Q_1 date of arrival: August 6–th 2011;
- Q_2 date of departure: August 11–th 2011;
- Q_3 the transfer she prefers: airplane;
- Q_4 the city from which she departs: Milano;
- Q_5 the activity she wants to do: bungee jumping;
- Q_6 the preferred date in which doing the activity: August 10–th 2011;
- Q_7 where she wants to go: San Francisco.

The FSE addresses the single–domain queries to the domain-specific service providers as follows:

- to the advertising service provider specific for hotels (e.g., Booking), the FSE addresses a single query composed by August 6–th 2011, August 11–th 2011, and San Francisco;
- to the content service provider specific for transport (e.g., Lufthansa website [15]), the FSE addresses a single query composed by August 6–th 2011, August 11–th 2011, airplane, Milano, and San Francisco;
- to the content service provider specific for activities (e.g., San Francisco Travel [19]), the FSE addresses a single query composed by bungee jumping, August 10–th 2011, and San Francisco;
- to the integrated content and advertising service provider (e.g., SanFrancisco.com [17]) and to the advertising service provider specific for restaurants (e.g., Best of San Francisco Guide [18]), the FSE addresses a single query composed by San Francisco.
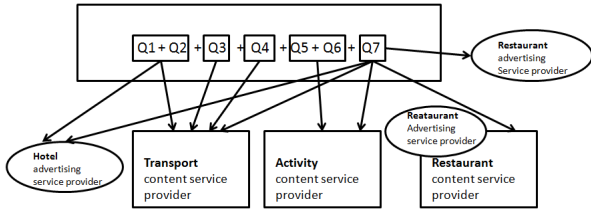
Figure 2: An example of interaction among actors.

4. A MICROECONOMIC MODEL FOR FEDERATED SEARCH ENGINES

The aim of this section is the design of a suitable microeconomic model to determine the appropriate payment for every involved actor. As discussed in the previous section, the proposed business model for the FSE involves a number of basic business models described in Section 2.1 and microeconomic models described in Section 2.2. In addition, specifically for the FSE, we need a novel microeconomic model to capture the integration of the advertisement and to find the appropriate redistribution. This model must determine the optimal payments between the FSE and all the advertising service providers.

We resort to auction models with redistribution schemes. Our model extends the models described in Section 2.2. In order to integrate different sources of ads and make it in efficient way, the FSE needs to have all the information about the ads (i.e., the qualities and the values). Otherwise, the FSE would not be in the position to target at best the ads for the user and could not extract the maximum revenue from the advertisement (e.g., without information on the ads, the FSE could display ads whose click probabilities are not the largest ones). The resulting auction model (where the FSE is an auctioneer and the advertising service providers are the bidders) prescribes that the bidders submit the qualities of the ads in addition to the values.

Model definition. The formal microeconomic model is a tuple:

$$\mathcal{M} = \langle S, A, X, \langle \Theta_s \rangle_{s \in S}, f, \langle p_s \rangle_{s \in S} \rangle$$

where $S = s_1, \ldots, s_n$ is the set of advertising service providers (included integrated content and advertising service providers), $A = a_1, \ldots, a_m$ is the set of advertisers; X is the set of possible allocations, where an allocation x defines which ads are displayed by the FSE and in which order; Θ_s is the set of possible combination of value and quality, one for each ad, that advertising service provider s can communicate to the FSE; $f : (\Theta_{s_1} \times \cdots \times \Theta_{s_n}) \to X$ is the social choice function, i.e., given the values and qualities communicated by all the advertising service providers, the social choice function determines the allocation; p_s is the payment rule of advertising service provider s. In addition, we define t_s as the minimal revenue the FSE must give to actor s according to a given contract and r_s as the redistributed revenue for s. The total revenue of s is defined as $\theta_s - p_a + r_s$.

Requirements. We state the microeconomic requirements. Some of these are the classical ones [11].

Individual rationality: the monetary revenue expected by each actor (except the FSE) is non–negative.

Weak budget balance: the monetary revenue expected by the FSE is non–negative.

Incentive compatibility: no actor can gain more by misre-

porting its true valuation (in the case of advertising, qualities and the values of their ads).

While the first two requirements are obvious, the third one may be not. This last requirement is necessary for the stability of the market. In absence of this requirement, the bids of the actors can fluctuate during time, reducing the revenue for the actors. An additional classical microeconomic requirement, needed to extract the maximum revenue from the auction, is:

Allocative efficiency: the chosen allocation maximizes the cumulative expected revenue.

Due to the business model defined in the previous section, we need to introduce two additional requirements:

Redistribution: the revenue must be shared in some way over the actors without violating the above requirements.

Flexibility: the microeconomic model must work with potentially different constraints due to different contracts.

Model design. In order to satisfy the above requirements, we design $\mathcal{M}$ (in terms of f and p_s) as an extension of the VCG mechanism (it is the unique mechanism satisfying the requirements [16]). It is composed of two computational phases: in the first phase, a VCG is solved to define the list of ads to display and the payments without redistribution; in the second phase, the redistribution is computed.

Phase 1. The social choice function is defined as:

$$f(\theta) = \arg\max_x \sum_s \theta_s(x)$$

the VCG payments are defined as:

$$p_s(\theta) = \sum_{j \neq s} \theta_j(f(\theta_{-s})) - \sum_{j \neq s} \theta_j(x).$$

where $\theta = (\theta_{s_1}, \ldots, \theta_{s_n})$ is the set of communicated combinations of values and qualities and θ_{-s} is the set of communicated combinations of values and qualities except the ones communicated by service provider s.

We define in addition:

$$VCG(\theta) = \sum_s p_s(\theta).$$

Phase 2. In the case the FSE is a private entity and the redistribution must be the minimum, each actor receives exactly $r_s = t_s$. In the case the FSE is a no–profit entity, the computation of the redistributions is more complicated. (Redistribution cannot be defined arbitrarily, otherwise the above requirements would not be satisfied.) Call π a priority over the actors. The redistribution to s is:

$$r_s^\pi(\theta_{-s}) = \min_{\theta_s' \in \Theta_s} \left\{ VCG(\theta_s', \theta_{-s}) - \sum_{\pi(i) < s} r_i^\pi(\theta_i) - \sum_{\pi(i) > s} t_i \right\}$$

where θ_s' is a possible combination of values and qualities communicated by service provider s, $\pi(i) < s$ denotes the bidders i that precede s in π, and $\pi(i) > s$ denotes the bidders i that follow s in π. This redistribution scheme assures that the redistributed revenue is the maximum one, as shown in [13].

Properties. First, we need to report an impossibility result. The above auction model determines payments independently of the actual ads clicked by the user and these payments cannot be imposed by a pay–per–click scheme. As shown in [5, 6], the pay–per–click scheme cannot be used without violating incentive compatibility when different ad sources are integrated. The unique payment scheme we can

use is pay–per–impression. The redistribution scheme is guaranteed to be undominated, i.e., it is not possible to redistribute more revenue without violating the incentive compatibility property. We experimentally evaluate the average percentage of the redistributed revenue with the *Webscope A3* Yahoo! dataset. This value is about 70%. That is, in many practical cases it is not possible to redistribute all the revenue to the service providers. We will experimentally evaluate the revenue of the FSE with Yahoo! dataset in future work.

Example. We consider the running example described in Section 3 and we suppose that the integrated content and advertising service provider s_1 communicates the ad a_1 to the FSE, the advertising service provider for restaurants s_2 communicates ad a_2, and the advertising service provider for hotels s_3 communicates the ad a_3. The products between the values and the qualities of these ads follow: $\theta_{s_1} = 0.4$, $\theta_{s_2} = 0.45$, and $\theta_{s_3} = 0.2$. Moreover, we suppose that the FSE has a contract with each content service provider that establishes how much the FSE has to pay them for each query. The FSE has to pay the content service provider for transport $0.03 per query, and the one for activity $0.05 per query. Since the integrated content and advertising service provider must not have incentives to communicate only its organic search results, i.e., it is not motivated to behave as a content service provider, the FSE has to guarantee it a revenue of at least $0.05, i.e., max$\{\$0.03, \$0.05\}$. Thus, we find out that the sum of the expected payments received by the FSE is $0.4 per query, and that the surplus it can redistribute is equal to $0.32. We consider that the priority π is: the integrated content and advertising service provider for restaurants, the advertising service provider for restaurants, and the advertising service provider for hotels. Applying the redistribution mechanism we propose, the FSE redistributes only to the integrated content provider a value of $0.12.

5. CONCLUSIONS AND FUTURE WORKS

A new search computing paradigm, called federated search engines, consists in the integration of heterogeneous domain–specific service providers able to scour the deep Web and find information that the current general–purpose search engines are not able to discover. In the literature, a lot of works that deal with this paradigm recently appeared, but none of these studies is about revenue mechanisms suitable to FSEs. Designing such a revenue mechanism is an issue of paramount importance for the success of FSE paradigm. In this paper we have initially analyzed existing revenue mechanisms. In particular, we have described the business models, that specify who pays and when, and the microeconomics models, that specify how the optimal payment can be computed, that are adopted by the existing mechanisms. We point out that the FSE requires a new and complex business model and none of the known microeconomics models can be used in our scenario. Due to this, we have faced the problem of designing a new heterogeneous and flexible revenue mechanism for FSE that satisfies some desirable properties and that allows for redistribution of FSE's surplus.

Our next step consists in doing an experimental evaluation of the mechanism we have proposed in this paper. This experimental evaluation involves the presence of humans that interact with a demo of the FSE that integrates organic search results as described in the SeCo project, and that uses the revenue mechanism proposed in this paper for

the merged ads list. Moreover we would like to understand how to target at best the displayed ads for the specific user that has done the search. To reach this goal we have to accomplish two tasks. First, we have to deeply investigate the different ways in which ads belonging to different lists can be merged together depending on the search that has been done, i.e., the problem is which ads to select. Second, we have to define multiple different user models, to capture the interest level of the user in the ads and how she looks at them, i.e., the problem is to understand how to display at best the ads given the model of the user that has made the search. There is also an interesting future search direction related to integrated content and advertising service providers and content service providers that display ads provided by an advertising service provider. In this case, the organic search results and the ads list can contain the same links. Obviously, displaying such links in both the results lists is not the most efficient solution. We need to design a new business model that defines the best strategy that these service providers must adopt in such a situation.

6. REFERENCES

[1] AdSense. http://www.google.com/adsense/.
[2] AdWords. http://www.adwords.google.it/.
[3] Bing. http://www.bing.com/.
[4] D. Braga, S. Ceri, F. Daniel, and D. Martinenghi. Mashing up search services. *IEEE Internet Computing*, 12(5):16–23, 2008.
[5] S. Ceppi and N. Gatti. An automated mechanism design approach for sponsored search auctions with federated search engine. In *AMEC*, pages 127–140, 2010.
[6] S. Ceppi, N. Gatti, and E. Gerding. Mechanism design for federated sponsored search auctions. In *AAAI*, 2011.
[7] Ecosia. http://ecosia.org/.
[8] Edreams. http://www.edreams.it/.
[9] Exalead. http://www.exalead.com/search/.
[10] Expedia. http://www.expedia.it/.
[11] D. Fudenberg and J. Tirole. *Game Theory*. 1991.
[12] Google. http://www.google.com/.
[13] M. Guo and V. Conitzer. Undominated VCG redistribution mechanisms. In *AAMAS*, pages 1039–1046, 2008.
[14] Ixquick. http://www.ixquick.com/.
[15] Lufthansa. http://www.lufthansa.com/.
[16] Y. Narahari, D. Garg, R. Narayanam, and H. Prakash. *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*. Springer, February 2009.
[17] Restaurantadv. http://www.sanfrancisco.com/restaurants/.
[18] Restaurantint. http://bestofsanfrancisco.net/restaurants.htm.
[19] SanFrancisco. http://www.sanfrancisco.travel/.
[20] SeCo. http://www.search-computing.it/s.
[21] Tripadvisor. http://www.tripadvisor.it/.
[22] Trivago. http://www.trivago.it/.
[23] Vivisimo. http://www.sedoparking.com/vivisimo.it.
[24] Wikipedia. http://it.wikipedia.org/.
[25] Yahoo! http://www.yahoo.com/.
[26] Zillow. http://www.zillow.com/.