2016 IEEE International Conference on Computational Science and Engineering, IEEE International Conference on Embedded and Ubiquitous Computing, and International Symposium on Distributed Computing and Applications to Business, Engineering and Science

Personalizing Information Retrieval using an Extension of a Dung Argumentation

Youssef Meguebli, Bich-Liên Doan, Fabrice Popineau
LRI, CENTRALESUPELEC
3 rue Joliot-Curie, Gif-Sur-Yvette, France
{first_name.last_name}@centralesupelec.fr

***Abstract*—In this paper, we propose an architecture for personalizing information retrieval (IR), exploiting the interactions between the user and the social network. We use an extension of a Dung argumentation framework to show how the precision of the personalized information retrieval system could be improved. We use also social media and search history to define the user-profile which is represented by a restriction of a Description Logic.**

Keywords— Information retrieval, opinions engineering, Dung argumentation framework, User-profile model, social media.

I. INTRODUCTION

Web personalization has become one of the most promising approaches to handle the main problem of information retrieval systems: the difficulty to determine the most relevant resources for a particular user, given a general query, with the explosive growth of resources in the web. Also, the Web information changes dynamically, so the search engine cannot be sure to update in time. According to the evaluation of experts, the rate of relevant results the current major search engines return is less than 45%. The aim of personalized IR is to model the user and to integrate her profile in the information retrieval process in order to provide Web information that matches with her personal interests. Despite a great deal of research, a number of challenges still exist, like managing the growth of the user interests and preferences over time.

Online social networks are quickly emerging as one of the greatest popular services on the Web. These systems are able to capture a weighty portion of Web users: for example at march 2016, Facebook counts more than one billion and a half monthly active users and one billion daily active users active on average during the same month.[1]

Current information retrieval models lack two essential properties needed to get the adequate answer to the challenges:

- lack of models to index appropriately the opinions: classic information retrieval systems index opinions of the users and sources (documents, video, …) but they neglect to represent the value judgment of the opinions about the sources.

[1] http://expandedramblings.com/index.php/by-the-numbers-17-amazing-facebook-stats/

978-1-5090-3593-9/16 $31.00 © 2016 IEEE
DOI 10.1109/CSE-EUC-DCABES.2016.261

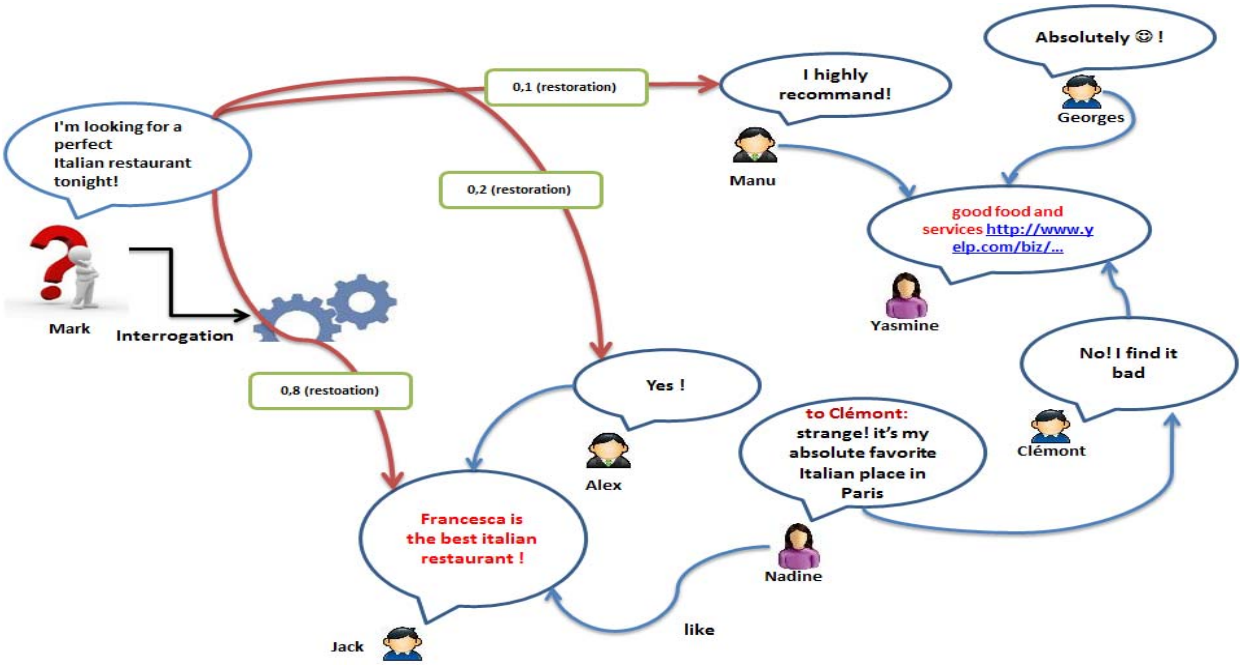

Figure 1

- the lack of mechanisms for judging the adequacy and the relevance of an opinion based on user profile preference, interest and level of trust related to a specific field

Consider a simple scenario (figure 1) to highlight the importance of opinions in user's social networks. With this scenario, we intent to show that it is desirable to exploit the opinions and relations between them in order to determine the most relevant information for a user. To better understand our goals, we propose the following example. We suppose that our user *is looking for an appropriate Italian restaurant for dinner*. Many factors are very important to get the most precise answer to this query. We start with a set of sources that provide assertions about some topic. Among the sources, we will find user's acquaintances connected through social network, micro blogs or other anonymous persons or entities that offer and spread data via a website. For each of these assertions, it is possible to have multiple opinions; those opinions can be positive, negative or neutral. Moreover, we can find opinions that criticize other opinions and others who appreciate or support other opinions. This is particularly the case on social networks.

Another criterion that we want to focus on for the selection and ranking of the results is surely the level of trust for any source that can differ from a domain to another. For our example, the degree of trust for each user depends on restaurant domain.

In figure 1, we present a simple scenario and the ideal solution to meet the specific need. Consider a user named Mark looking for the best Italian restaurant. The system has indexed information about two restaurants published by Jack and Yasmine. The assertion published by Jack was appreciated by two positive opinions from Nadine and Alex notably by a like action from Nadine and a comment from Alex. The second assertion shared by Yasmine was both praised and criticized by her friends. Also, there are some opinions related to other opinions: For example, Clemont criticized Yasmine assertion and he has also been criticized by Nadine. For each friend, Mark has a trust level on catering field.

Opinion mining is a growing research area both at information retrieval and natural language processing communities. Often, customers use beliefs, judgments and opinions that people may express in social networks blogs and micro-blogs, reviews, audio and video sharing website to decide and judge the information suitable to their interest and preference. In this paper, we present a smart information retrieval architectural able to fix the most relevant results based mainly to (i) the opinions related to every concept or object and (ii) the trust property for every source. In this work, we suppose that our user is looking for an appropriate restaurant for dinner. In our architecture we raise more interest to the source and it will be an important factor for the query process. The source can be, in our context, a person, web site or also an anonymous author or published of an information or opinion. Among the sources, we will find user's acquaintances connected through social

network, micro blogs or other anonymous persons or entities that offer and spread data via a website.

For each of these assertions, it is possible to have multiple opinions; those opinions can be positive, negative or neutral. Moreover, we can find opinions that criticize other opinions and others which appreciate or support other opinions. This is particularly the case on social networks, Photo and videos sites. Another criterion that we want to focus on for the selection and ranking of the results is surely the level of trust for any source that can differ from a domain to another.

For our example, the degree of trust for each user depends on restaurant domain. Current information retrieval models miss essential property needed to get the adequate answer in this scenario. The lack of mechanisms for judging the adequacy and the relevance of opinion based on social context of the user profile preference mainly the level of trust for every source and related to a specific field.

In this Work, we present the adequate components to tackle these problems. We aim to determine the most suitable restaurant from our user profile preferences, interests and based on opinions collected in our social network. We will not tackle the problem of extracting and formalizing opinions out of other user's interaction with the social network. This was investigated in [1] which proposes a semi-supervised topic modeling approach to sentiment analysis. Finding relevant information is obviously important, but there is (allegedly) a tendency for people to only look at information they find "comfortable": opinions and reviews forwarded or recommended by friends or appearing in familiar and friendly venues. Information retrieval technology can and should also be used to remind people that there are other perspectives that they might want to consider, both for topics they're already familiar with and for new topics they are considering. [2] is an example of system that attempts to do that by showing multiple perspectives for "controversial" topics. Our contribution with this paper is a new architecture that can handle some classical problems. We present how to model the user-profile, and how we can answer a query with the most acceptable opinions using a Dung argumentation framework [3].

II. ARCHITECTURE

The goal of our system is to be both personalized and personal, whereas usual IR systems that can be personalized but that are still global. So our system builds an index for each user. Each user can define his own sources among blogs, social networks, and RSS feeds and so on.

The following stages describe our architecture from an operational point of view.

1- Stage 1: the user declares his domains of interest relatively to the reference ontology and chooses a set of sources. The user may tag sources to describe their domain of relevance, or the trust level he has in each source about their domain.

2- Stage 2: sources push and index information continuously to the user. In parallel, the system build initial user-profile model and start the process of evolution over time.

3- Stage 3: the user queries the index

Our architecture is composed by three main components notable user-profile model, query processing and index processing. In this paper, we interest mainly by the query process which present our contribution. Indeed we exploit an extension of argumentation framework to fix the most

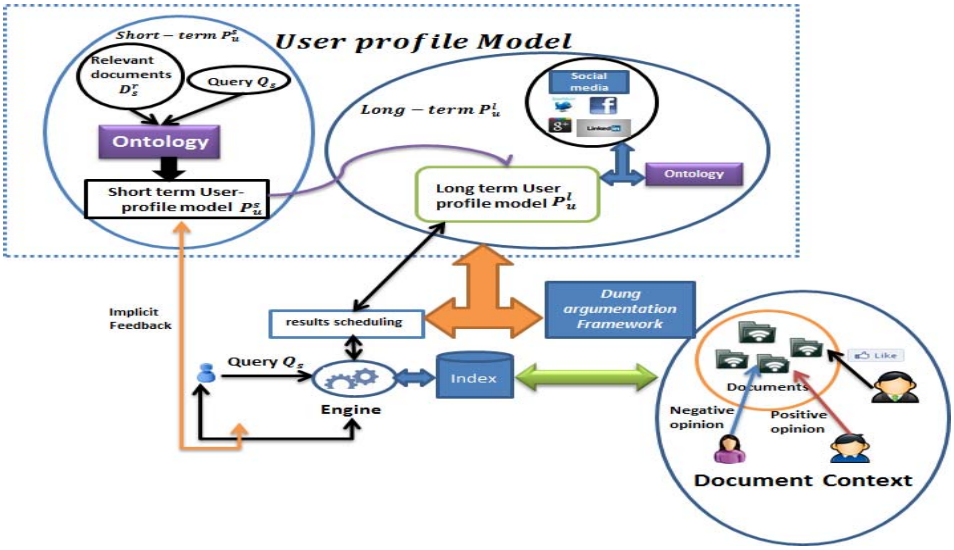

Figure 2 Overview architecture

relevant data. The main idea of this work is to determine the context of each object (document, video, image…) principally the opinions related to one object. The current information retrieval uses a statistical technique to fix the most relevant without dealing the relation between the different opinions and also they consider the different opinions with the same trust level. In our case we enrich each opinion by a trust level related to the author and the domain. We also deal the relation between each opinion to fix the most relevant opinions for each object. For each user query, we consider the available opinions in order to determine the most relevant answers according to the Dung argumentation framework. The model that we propose in this paper can be distinguished by the ability to use the Dung argumentation framework to retrieve the most acceptable opinions for each resource; this will allow sorting the different resources based on the acceptance value of each opinion and on the trust level of each author.

III. USER-PROFILE MODEL

User modeling defines the creation and evolution processes of a set of properties about all aspects of the user, which are pertinent to the adaptation of user interactions. For instance, this can include personal context notably the interests, level of expertise, preferences and information about social context to infer some properties from these social relations (friends). In this work, we exploit mainly the social context to fix the trust level related to each domain for one source. Indeed, in our user-profile model we define the trust related to one domain and one source.

In our user profile model, we make a distinction between the short-term user profile which is determined after each session, a session being defined by a set of consecutive queries in the same domain, and the long-term user profile which is determined by social, personal and spatio-temporal contexts over longer period of time**.**

Many structures and paradigms have been proposed in the academic literature for the representation of the knowledge and information concerning users. User profile model can be represented as a class vectors or a set of keyword vectors [5], a set of concepts [6] or an instance of predefined ontology [7,8]. In our architecture we use a restriction of description logic [9] to represent our user-profile model for mainly two reasons:

- it offers a highly expressive ground for describing the user profile model;

- it allows the use of inference rules to deduce new concepts from other concepts.

The short term user-profile model corresponds to a set of concepts and a degree of interest, which are deduced after each session. The session is a set of consecutive queries $\{q_1, q_2, \ldots . q_n\}$ related to the same topic. We suppose that we have a limited number of topic (domains) that interests our user. We will not be interested here in the techniques and mechanisms to delimit the bound of the session. The short term user-profile model is represented by a set of concepts or instances, each one paired up with its level of interest:

$$P_u^s = \{(C_1, value)(C_2, value) \ldots .,\quad (C_n, value)\}$$

Where u is the user, s denotes short and $(C_1, C_2 \ldots . C_n)$ are the concepts and sub-concepts related to the main domain, and *value* is the level of interest corresponds to each concept.

The long term user-profile model denoted by $P_u^l = \{P_{c1}, P_{c2}, \ldots \ldots, P_{cn}\}$ is composed of a set of user-profiles. Each user-profile is specified for one topic or domain. For

instance, we can have a long term user-profile model for sport topic and another for politic topic. The user can define a priori her domains of interests and the system detect implicitly the other fields of interest based on the way the user interacts with her sources of interest. For any P_{ci} we model the personal, spatio-temporal and social context. The personal, social and spatio-temporal of these profiles are deduced mainly from the user interaction with social media and search history. The user profiles is created based on explicit and implicit actions: the user can define some properties (age, interest, level of interest, trust degree for each source…) but the system can also deduce this information from sources. For example, the social network profiles have generally an acceptable degree of credibility. So, we can define the user-profile on social networks as a source (by default is just a medium between our system and sources) and we can give it an acceptable trust level. So based to this source we can extract relevant information related to our user. After the construction step, our user can validate or change some properties of this model. Next, we start the evolution process of our user-profile model based essentially on her/his actions with social media and search history that are represented by the short term user-profile. We use also the expressive power of the Description Logic and the inference service to deduce new properties. For instance, we can exploit the subsumption inference to deduce new interest (sub-interests) from other interests. We present here a simple algorithm which extracts data from a set of sources behind Facebook social Networks.

IV. QUERY PROCESS

We denote $< o_1, o_2, \dots o_n >$ a set of opinions related to an assertion (concept). Each Opinion O_i is ranked with a value between [-2,2] referring to {very negative, negative, neutral, positive, very positive} appreciation. Opinions can attack or approve an assertion, or themselves. Each opinion has an author and also a level of trust related to a specific domain [10].

So we have a set of tuples of the form $< O_i, t_i >$ for each assertion where O_i is the opinion and t_i is the trust level of the source. The trust level is deduced from social context user-profile which represents the trustiness for each source to a specific domain. A problem to solve is how to determine the most relevant information based on different opinions. Each opinion has a level of trust related to the domain and the author. A Dung argumentation framework is able to solve this problem since it includes a set of arguments and a binary relation between arguments which is typically based on a notion of attack.

A key issue is the interaction between arguments, notably the way arguments are generated, how they interact and how to evaluate them, and also how determine the most acceptable arguments. An argument can be seen as a reason for supporting some claim. Conflicts between arguments arise for example when the claim or the reason for supporting it is contradicted by another argument.

Evaluation then aims at selecting the most acceptable arguments. The semantics given by Dung for acceptability define (with respect to each argument system) one or several acceptable sets of arguments called extensions. All the Dung's semantics rely upon the notion of an admissible set, but this is not shown in the original definitions [3].

In this section we present how we can use this framework to fix the most acceptable opinions related to each instance of concept and consequently help the actual information retrieval system to answer more precisely to the user queries.

Classic information retrieval systems often give numeric scores to resources and then rank them based on the scores in order to make recommendations to users. We can mention some information retrieval techniques used such as the space model, the probabilistic model and the inference network model.

In our architecture we exploit the trust level of each source to determine the relevant results. For instance, we index the instances (or objects) like 'restaurant1' and we index also the opinions related to each object. Our idea is to infer a trust level for each data given the trust level of the source.

Indeed, for each opinion having a trust level we simply pass this value to the object. For instance, if Paul said that Restaurant "Aldo" is a good restaurant and our user-profile has a trust level of +2 for this source towards restaurants, then "Restaurant Aldo is a good restaurant" is endorsed with +2. We use opinions based on an extension of a Dung argumentation framework to first determine the opinions related to our query (restaurant in our example), second get the most acceptable opinions for each resources, and finally determine the most relevant data based on the degree of each opinion ranging from 'very negative' (or -2) to 'very positive' (+2) and trust level for each author.

We use a cascading system that first select the data which responds both to the different contexts and search history. For some queries we get the set of opinions related to each concept or instance. Based on an extended Dung argumentation framework we determine the most acceptable opinions according to the level of trust of each source and according to the different opinions emitted about the same subject.

Indeed, in this architecture we extend Dung's Argumentation Framework (AF) into a Trust-Based Argumentation Framework (TAF), by allowing taking into account the trust level of each author to extract the acceptable opinions. An opinion will qualify to be acceptable if it is endorsed with a trust level greater than 0. This information will be represented as follows:

Definition : A trust-Based argumentation framework (TAF) is 4-Tuple:

TAF = <A, R,V,Trust> Where ***A*** and ***R*** are the same as for a standard argumentation framework, ***V*** is a non-empty set

of values that contains the different trust value of each opinions authors and **Trust** is a function which maps from elements of ***A*** to elements of ***V***.

We say that an argument ***A*** relates to value ***V*** if accepting ***A*** promotes or defends ***V***: the value in question is given by ***Trust(A)***. For every $A \in R$, $Trust(A) \in V$. Our purpose in extending the AF is to allow us to distinguish between one argument attacking another, by taking in consideration the trust level of each author. Based on this extension Dung's AF we are able to determine the most acceptable opinions for each concept.

Whether an opinion 'a' attack other opinion' b' and the author of 'a' has a trust level more important than the author of 'b'. We automatically delete the first opinion 'a'. Consequently, in our previous example, we initially select the set of restaurants that fit to our user-profile (based on our model which represents the different contexts like the spatio-temporal one and the personal one) like the most actual personalized information retrieval.

Secondly, we determine the different opinions related to each restaurant that was selected after this first filter and we use our Trust-based argumentation framework to return the most acceptable opinions based on social context of our user-profile model (especially the level trust related to each friend).

Finally, we sort based on opinions by calculating with a simple formula the value which depend on the trust level of each author and the degree of every opinions (which depend from -2 to 2). So the data will be sorted and returned based both on user profile interest, preferences (context of our user profile) and the opinions related to the query.

V. CONCLUSION

To build efficient information retrieval system we must understand and analyze the complexities of the human information seeking. In this paper, we propose a smart architecture of information retrieval system that take into account the opinions notably by classify and fix the most acceptable opinions based to trust property. We have also proposed a user-profile model which based to search history and social media which represent the personal, spatial-temporal and social context. There are many directions in which this work can be extended. First, we plan to further analyze the user-profile model by exploring the social context (especially the trust level related to each source for every domain interested our user) based on social media such social networks and micro-blogs. In this context, we plan to investigate how we can extract data from different sources, and how we will represent this data and also the inferences rules that allow inferring other relevant data. We also plan to further investigate the compactness aspect of the Dung argumentation framework. In effect, the argumentation framework can be exploited and it is a promising technique to determine the most acceptable opinions for each source and consequently help to get the most relevant results for each query.

REFERENCES

1. Lu, Y. , Zhai, Ch. Opinion integration through semi-supervised topic modelling, In Proc. of WWW 2008,. , Bejing, 2008, pp. 121-130.
2. Kacimi, M.,Gamper, J. "Diversifying Search Results of Controversial Queries". Proceedings on CIKM2011, the 20th ACM Conference on Information and Knowledge Management. Glasgow-Scotland, 2011.
3. Dung, P.M. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artificial intelligence, 77(2), ,1995, 321-358.
4. Kofod-Petersen, A., Aamodt, A. Case-Based Situation Assessment in a Mobile Context-Aware System. In A. KrÄuger and R. Malaka, editors, Artificial Intelligence in Mobile Systems (AIMS), pages,USA, 2003, 41-49.
5. Tamine-Lechani, L. , Boughanem, M., Zemirli, N. Personalized document ranking: exploiting evidence from multiple user interests for profiling and retrieval. In Journal of Digital Information Management, 2008.
6. F. Liu, C. Yu, and W. Meng.Personalized web search for improving retrieval effectiveness. IEEE Transactions on Knowledge and Data Engineering, 16(1), 2004, 28–40.
7. Gauch, S., Chaffee, J.,Pretschner, A. Ontology-based personalized search and browsing. Web Intelligent and Agent Sys., 2003, 219–234.
8. Sieg, A., Mobasher, B. ,Burke.Web, R. search personalization with ontological user profiles. InCIKM'07: Proceedings of the ACM Conference on information and knowledge management, pages525–534, New York, NY, USA, 2007.
9. Baader, F., Calvanese D., McGuiness,D., Nardi,D., Patel-Schneider, P.F. editors. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2003.
10. Atencia, M., Euzenat, J., Pirro, G., Rousset, M.C. Alignment-based trust for resource finding in semantic P2P networks. Proceedings on 10th International Semantic Web Conference (ISWC), Germany,2011.