AMBIT: Semantic Engine Foundations for Knowledge Management in Context-dependent Applications

Riccardo Martoglia

FIM Department, University of Modena and Reggio Emilia, I-41125 Modena, Italy,
E-mail: `riccardo.martoglia@unimore.it`

Abstract **– Context-aware application and services proposing potentially useful information to users are more and more widespread; however, their actual usefulness is often limited by the "syntactical" notion of context they adopt. The recently started AMBIT project aims to provide a general software architecture for developing semantic-based context-aware tools in a number of vertical case study applications. In this paper, we focus on the knowledge management foundations we are laying for the Semantic Engine of the AMBIT architecture. The proposed semantic analysis and similarity techniques: (a) exploit the textual information deeply characterizing both users and the information to be retrieved; (b) overcome the limits of syntactic methods by leveraging on the strengths of both classic information retrieval and knowledge-based analysis and classification, ultimately proposing information relevant to the user interests. The experimental evaluation of a preliminary implementation in an actual "cultural territorial enhancement" scenario already shows promising results.**

Keywords **– context-aware applications; information retrieval; text analysis; semantic knowledge and similarity.**

1. Introduction

Nowadays, we are constantly supported by ICT systems and applications that exploit ubiquitous services supporting different kinds of human activities. However, the availability of a large number of services can turn out to be confusing rather than useful, since the users are often overwhelmed by the large number of "proposals" which they are generally not able to consider thoroughly to find what they really need. To overcome this problem, many researchers have proposed to develop new applications with (or to incorporate in existing applications) context-awareness capabilities ([2, 4]). A context-aware application is one that "knows" the context in which the client is operating and possibly also the profile/characteristics of the user who is enjoying the corresponding service(s). Clearly, such knowledge must be gathered (often under real-time constraints), stored in well-organized fast-access data and information systems, and effectively exploited with the goal of delivering "personalized"

high-quality *context-dependent* services. This is far from simple; the main limitations of existing efforts lie in the limited notion of context they adopt, and especially in the almost complete absence of any attempt to model the semantics of the context.

This is the challenging scenario of the recently started AMBIT (Algorithms and Models for Building context-dependent Information delivery Tools) project[1] [5] a regional project co-funded by Fondazione Cassa di Risparmio di Modena and managed by the Softech-ICT research center. The main goal of the project is to study and implement a prototype software architecture for the development of *context-dependent applications and systems*, i.e., tools that provide users with services that are fully customized according to the context in which they operate. Preliminary steps will be the study of models, algorithms and data structures for the representation and manipulation of contexts. AMBIT will study and implement a very broad idea of context, including (among others) the modeling of the external environment, the users' profile and the history of the actions performed by them.

The AMBIT software platform will eventually provide an API that can be personalized for the development of a number of vertical context-dependent applications. Several case studies have been identified by the project industrial partners; one of the main application scenarios, and the one which will be the reference in this paper, is the "cultural territorial enhancement" one: through both on-demand and proactive services, users of specific applications (including mobile ones) are empowered with precious "suggestions" pointing to the information (e.g. territorial activities, typical products descriptions, tourism information, etc.) which is the most relevant with respect to their profile and needs. A very simple example could be the notification of an event which is geographically close to the location of the user (say, a country fair with local farm exhibitors), which falls under the interests associated with his/her profile (e.g., gardening enthusiast). Another example could be the monitoring of tourists interests (e.g. through a dedicated mobile app where tourists could browse information on Emilia-Romagna typical products) and, based on their favorite browsed pages and/or on explicit queries asking for specific information/topics, the retrieval of the pages that best capture their interest.

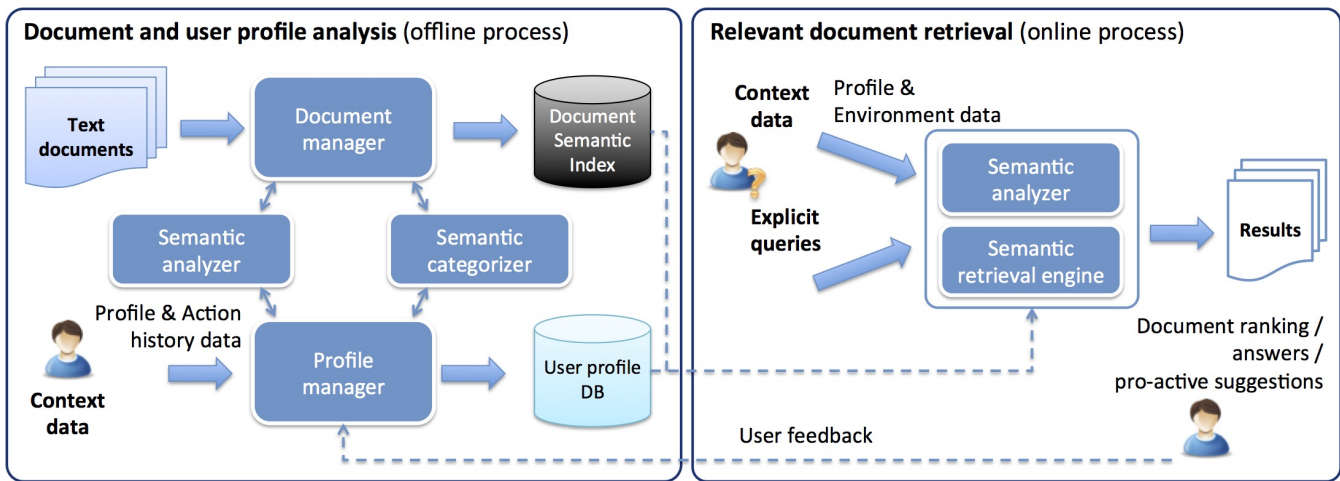[1] http://www.agentgroup.unimore.it/ambit/

Figure 1. An overview of the AMBIT-powered Semantic Engine for information management

In order to achieve the AMBIT goals, several studies on complementary techniques and research fields will have to be performed. One of the most crucial among them, the key to provide "intelligent" suggestions and answers to users, is certainly to have powerful ways of managing available information and knowledge. In this paper, we focus on the foundations we are laying for the **Semantic Engine** of the AMBIT architecture, and, in particular, for its *knowledge management techniques* that are indeed one of the most challenging aspects of the AMBIT project and should be ultimately able, together with other AMBIT results, to deliver high-quality context-dependent information. The techniques we propose:

- take advantage of textual information, certainly the primary component of the documents that should be presented / suggested to users, and also one of the major information characterizing user profiles (think, for instance, to the contents of their browsing history, to the description of their interests, and so on);

- are completely flexible and designed to be easily applicable to the territorial enhancement scenario considered in this paper, but also to all the application scenarios involved in AMBIT (which also include, among others, context-aware advertising, smart help-desk problem solving, etc.).

More specifically, Figure 1 shows an overview of the main processes (and the related modules) of the semantic engine which will allow AMBIT-powered systems and applications to:

1. manage document and profile information (**Document and user profile analysis**, left part of Figure). This is done by extracting and indexing the associated semantics by means of ad-hoc *semantic text processing* and text classification techniques (described in Section 2), also exploiting external knowledge sources;

2. provide useful answers/proactive suggestions to the user (**Relevant document retrieval**, right part of Figure), by

retrieving the most relevant documents w.r.t. the user profile and/or query. This is achieved thanks to novel and specifically devised *semantic similarity* techniques (detailed in Section 3).

Section 4 shows preliminary but already promising results and the good effectiveness of the proposed techniques, by means of an experimental evaluation done on a small-scale actual territorial enhancement scenario. Finally, Section 5 concludes the paper also by briefly analyzing related works.

2. Document and user profile analysis

Document analysis. In this offline process, which is propaedeutic to the online document retrieval process (Section 3), the available documents are processed and the information which will be required in the actual retrieval is extracted, stored and indexed in an ad-hoc *Document DB* by a *Document manager* module (see left part of Figure 1). The input information are the text documents relevant to the specific scenario instantiation, including available web pages, product and service descriptions, and so on. For instance, in our territorial enhancement use case, these include descriptions of fairs and events which have been or will be held in the area, descriptions of typical products, details on forthcoming initiatives and activities, information on touristic points of interest, etc.

Since existing packages do not allow sufficient configuration and extension options, we preferred to design a custom-made *Semantic analyzer* tailored to the AMBIT environment. The analyzer performs several steps which are needed in order to extract the contents (and meaning) of the processed information, including: *Tokenization*, the terms of the different sections are identified and punctuation is removed; *Stemming*, the tokens are "normalized" and "stemmed", i.e., terms are reduced to their base form (managing plurals and inflections); *POS (Part of Speech) Tagging*, the tokens are "tagged" with Part of Speech tags (i.e., nouns, verbs, ...); *Composite term identification*, pos-

a)

TERM	SYNS	IS_A	DEFINITION	IDF	DOC_LIST
Expo	Exhibition, Exposition	Collection, aggregation	A collection of things (goods or works of art etc.) for public display	2.455	['D07542-3', …]
Parmesan		Cheese	Hard dry sharp-flavored Italian cheese; often grated	7.457	['D03522-3', 'D08654-2', …]

b)

CLASS	IS_A	DEFINITION	DOC_LIST
Automotive equipment	manufacturing& engineering	Companies that produce components for automobiles	['D04342', …]
Viniculture	agriculture	Production of wines from the vines to the finished products	['D03265', …]

c)

DOC	TERM	TF	W
D00001-1	Ferrari	0.545	0.977
D00002-1	ModenaTour	0.210	1.131

d)

DOC	CLASS	SCORE
D00001	Automotive equipment	0.645
D00002	Tour operator	0.410

Figure 2. Sample portions of the extracted Document Semantic Index: global view for terms (a) and classes (b), per-doc view for terms (c) and classes (d).

sible composite terms (such as "production area" or "wine tasting") are identified by means of a simple state machine and of POS tags information; *Filtering and enrichment*, terms are associated to additional information (such as definitions, synonyms, ...) extracted from the thesaurus (i.e., WordNet[2], [12]).

The extracted information will enable the retrieval of the most relevant information for the user in the online phase. For instance, by means of synonyms, documents about an "exhibition" will also be relevant to a query about an "expo".

Moreover, a *Semantic categorizer* processes text in order to tag each document with appropriate subject classes; we adopt the text-centric Media Topic NewsCodes taxonomies and vocabularies provided by IPTC[3], a well-known taxonomy offering a very good level of detail and coverage of the AMBIT topics. Each class tag has a score (the higher the score the more relevant the class is for the document). For instance, a document about the typical "Lambrusco" wine will presumably have "viniculture" among its highest scoring associated tags.

By applying batch document analysis to the document collection, the semantic index is automatically generated. Conceptually, it consists of a *global view* (all terms/classes together with their occurrences and additional extracted data, see Figures 2-a and 2-b for an excerpt) and a *per-document view* (terms/classes occurrences in each document with their statistics, Figures 2-c and 2-d). In particular, *DOC_LIST* is the list of the documents IDs in which each term/class occurs. Each occurrence is also associated to a weight reflecting its importance and meaningfulness in the text (*SCORE* for classes and *W* for terms, corresponding to the TF/IDF [14] model used in information retrieval). As we will see, this will allow the similarity functions of the Semantic Engine to draw useful knowledge from both the semantic and the classic text retrieval worlds.

User profile analysis. The *User profile DB* is populated and updated by the *Profile manager* each time a user connects. In particular, user context data may include *Profile data*, i.e. personal data, likings, preferences, etc. explicitly submitted by the user, *Environment data*, e.g. location data as extracted from

GPS sensors, time of day, ambient information such as lighting, noise level, etc, and *Action history data*, i.e. information about past user actions. This last kind of data is the one we mainly focus on in this paper and is particularly crucial for the semantic engine: it may include, for instance, past accessed documents (e.g. browsed from an ad-hoc AMBIT app), past actions performed on some partner's website (e.g. about typical Modena products), and so on[4]. The intuition is that the documents from the user history can be analyzed in a similar way to the scenario documents, therefore exploiting all the power of the Semantic analyzer and categorizer in order to associate meaningful terms and classes to users (and thus enriching the user profile DB with information analogous to the one discussed for document analysis and shown in Figure 2). Due to their complexity, such analyses are performed offline and will be available for more accurately processing future requests from the same user.

3. Relevant document retrieval

This is the phase where users connect and receive the results which are relevant to their status or need (see right part of Figure 1). We encompass both the computation of proactive suggestions (based on context data) and the retrieval of "on-demand" personalized information to explicit user queries. For instance, a user could submit an explicit query about "Typical food stores" and the engine, also based on the user's past actions (i.e. browsing specific food descriptions) and environment, will produce a list of nearby stores which (s)he could find interesting, possibly personalized on the basis of its preferences/browsing history. However, the retrieval process could also work without any explicit user input. As an example, based on the current Profile/Environment data and on the user profile DB, the platform could detect that a user interested in sport cars is traveling by train and is reaching a stop near an international car fair; therefore, a proactive suggestion pointing to the fair web page would be pushed to his mobile device.

---

[2]http://wordnet.princeto n.edu/
[3]http://www.iptc.org/site/Home/

[4]The information may be directly available from the websites' logs or mobile application data or, where applicable, it may be indirectly derived by means of appropriate web tracking mechanisms ([1]).

In all cases, the *Semantic retrieval engine* module has to analyze the profile, environment and/or the query, access both the Document DB and the User profile DB data and produce a ranking of the available documents. In case of an explicit query, the ranking is directly presented to the user; in case of a proactive scenario, the suggestion for the document(s) on top of the ranking is sent to the user. We also plan to manage *feedback* on the relevance/usefulness of the received results; the profile manager will update the user profile information with typical requests and result feedbacks in order to dynamically modify preferences and, thus, avoid unwanted suggestions.

The computation of the document ranking is based on ad-hoc similarity metrics:

- the similarity $TextSim$ between the main terms of the available documents (and their sections) and those associated with the user profile (e.g. past navigated documents), possibly including explicit query terms;
- the similarity $ClassSim$ between the document and user classes;
- additional similarities on other aspects coming from profile and/or environment data, such as explicit preferences or likings, current time and location, etc.

As anticipated in the past sections, the need of effectively and efficiently computing similarities between the terms/classes characterizing user profiles and documents is crucial in AMBIT. We will now deepen the discussion of these two similarities, which are in the focus of the paper and are certainly key to the semantic understanding and satisfaction of the user query.

Text similarity ranking. $TextSim(U, D)$ quantifies, given a user profile U and a document D, the similarity of the user profile w.r.t. the document on the basis of their associated terms $t^U \in U$ and $t^D \in D$: In particular, the computation of $TextSim$ between a given profile U and each possible D (i.e., each available document in the semantic index) involves the following steps:

1. considering each term in U and finding the most similar term or terms available in D by exploiting a *term similarity formula $TSim$*;
2. inducing a **ranking** of the available documents (on the basis of $TextSim$), thus predicting which documents are relevant and which are not w.r.t. U.

Equation (1) shows the text similarity formula: the similarity is given by the sum (defined in (2)) of all term similarities between each term in U and each term in D maximizing the term similarity with the term in D:

$$TextSim(U, D) = \sum_{t_i^U \in U} TSim(t_i^U, t_{\bar{j}(i)}^D) \cdot w_i^U \cdot w_{\bar{j}(i)}^D \quad (1)$$

$$t_{\bar{j}(i)}^D = argmax_{t_j^D \in D}(TSim(t_i^U, t_j^D)) \quad (2)$$

where $w_i^U = tf_i^U \cdot idf_i$ and $w_{\bar{j}(i)}^D = tf_{\bar{j}(i)}^D \cdot idf_{\bar{j}(i)}$. In this way, each term contributes to the final similarity with a different weight. Moreover, the limitations of standard syntactycal techniques are overcome by computing $TSim$ by means of

Equation (3), which considers **synonyms** (as extracted in the semantic index) and **semantically related terms**:

$$TSim(t_i, t_j) = \begin{cases} 1, & \text{if } t_i = t_j \text{ or } t_i \, SYN \, t_j \\ r, & \text{if } t_i \, REL \, t_j \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Besides equal terms and synonyms, the formula provides a further case (REL) where the two terms are not equal or synonyms, nonetheless they are in some way related from a semantic point of view (i.e., broader/narrower terms etc.): such terms will contribute with a similarity of r, where $0 < r < 1$ is a user-defined fixed similarity value. REL is computed in real time by exploiting the relations between terms coming from the WordNet thesaurus and, more in detail, the method proposed in [8], a well established metric relying on the hypernym relations: for instance, "pasta" will result related with "dish" and "pizza".

Class similarity ranking. In addition to document terms, the classes associated by the semantic classifier can also significantly help in order to retrieve useful documents. This is obviously true if both a user profile and a document are strongly characterized by a common IPTC class (e.g. "Motor car racing"); however, also documents about Ferrari cars and tagged with a similar class "Formula One" would be of interest. This is achieved through $ClassSim(U, D)$ which quantifies, given a user profile U and a document D and in the same philosophy as $TextSim(U, D)$, the similarity of the user profile w.r.t. the document, in this case on the basis of their associated IPTC classes $c^U \in U$ and $c^D \in D$:

$$ClassSim(U, D) = \sum_{c_i \in U} CSim(c_i, c_{\bar{j}(i)}) \cdot s(c_i) \cdot s(c_{\bar{j}(i)})$$
$$(4)$$

$$c_{\bar{j}(i)} = argmax_{c_j \in D}(CSim(c_i, c_j)) \quad (5)$$

$$CSim(c_i, c_j) = \begin{cases} -ln\frac{len(c_i, c_j)}{2 \cdot H}, & \text{if } len(c_i, c_j) < Th \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

The similarity CSim (eq. (6)) between two classes c_i and c_j is derived from the hypernym metrics exploited for term similarity [8, 9]: it is computed as a score which is inversely proportional to the length of the shortest path connecting the two classes in the IPTC hierarchy; in case the length exceeds a configurable threshold Th, the similarity is null. H is a constant representing the maximum depth of the hierarchy (5 for IPTC).

Ranking fusion. The rankings $\tau_{text}, \tau_{class}$ induced by Eqs. (1) and (4), respectively, on the documents D given a profile U, are eventually fused in a final fused ranking $\hat{\tau}$ through a linear combination method [13], exploiting both terms and classes contributions:

$$s^{\hat{\tau}}(D) = \sum_{\tau \in \{\tau_{text}, \tau_{class}\}} \alpha_\tau \left(1 - \frac{\tau(D) - 1}{|\tau|}\right) \quad (7)$$

where $|\tau|$ is the length of the ranking and $\alpha_\tau >= 0$ is a preference weight (default is 1 for both rankings). Only documents which are part of both rankings will appear in the final

User req	Our Results			Typical retrieval baselines					
				Syntactic			Syntactic, no t.a.		
	Prec	Rec	F	Prec	Rec	F	Prec	Rec	F
U1	0.94	0.93	**0.94**	0.88	0.21	**0.34**	0.88	0.17	**0.29**
U2	0.90	0.93	**0.92**	0.90	0.87	**0.88**	0.24	0.34	**0.28**
U3	0.92	0.95	**0.93**	0.91	0.42	**0.58**	0.12	0.23	**0.16**
U4	0.94	0.89	**0.91**	0.78	0.31	**0.45**	0.67	0.23	**0.34**
U5	0.94	0.91	**0.92**	0.83	0.10	**0.18**	0.18	0.08	**0.11**
U6	0.92	0.93	**0.92**	0.82	0.24	**0.37**	0.12	0.13	**0.13**

Figure 3. Effectiveness analysis: precision, recall and F-measure (our results on the left, two baselines on the right).

ranking. Finally, note that, through α_{tau}, the two similarity formulas presented in this section can be combined in a completely flexible way, in order to make the Semantic Engine adaptable to the specific needs of different scenarios.

4. Experimental Evaluation

We will now present the preliminary results we obtained from an exploratory effectiveness evaluation we performed on a first prototype of the Semantic Engine, in the context of AMBIT. Together with the project partners, we considered a first simplified instantiation of the "cultural territorial enhancement" scenario; the document collection is composed of nearly 2000 documents about activities, tourism, food, exhibitions and fairs, manufactures, arts and events of the Modena and Emilia-Romagna area. Starting from this collection, which serves as a source of possible suggestions, we also considered different user profile requests ("users" in the following) simulating different kinds of interests/preferences. In this evaluation, being the effectiveness of the proposed analysis and similarity techniques the current focus, user information is strictly composed by an action history context (typically, 10-15 past navigated external documents witnessing their interests) and possibly by explicit query terms (2-8 queries). We will also assume a "stable" situation, where users and documents have been already automatically analyzed and their relevant terms and classes stored in the Semantic Index and User DB, respectively.

Among the considered users, we selected a set of 6 (U1-U6) as the most representative ones. For each one, we compared the output of the Semantic Engine with a "gold standard", i.e. relevant answers manually selected from the collection by expert project partners, and assessed precision, recall, and F-measure (Figure 3, left part). The results are compared with two baselines simulating the results offered by typical retrieval engines: a syntactic retrieval method ignoring synonyms, related terms and class information, and a syntactic method also ignoring text analysis (composite terms identification, stemming, etc.).

As we can see, the precision and recall levels achieved by the semantic engine are generally very satisfying: all users widely benefit from the proposed semantic features. Let us now analyze the results in detail. Typical terms and classes associated to U1 involve generic "tourism information": documents about
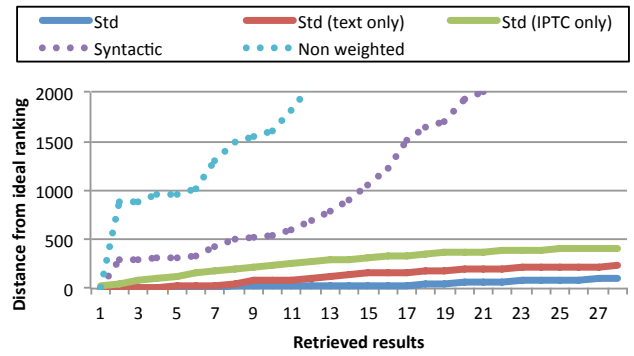

Figure 4. In-depth effectiveness analysis for U5: distance from optimal ranking

restaurants, hotels and tour operators are correctly retrieved by the engine due, for instance, to the similarity of the "tourism" "catering" and "accommodation" IPTC classes and of the contained terms (recall of 93%); on the other hand, they do not necessarily contain the same terms present in the user profile, thus syntactic techniques are not able to identify most of the relevant documents (recall of 21% or lower). Text analysis is also key to good results: for instance, the absence of stemming has a very negative impact on the second baseline recall for all users. Moreover, U2 and U3 are characterized by a high number of composite terms: for instance, U2 is mainly interested in "wine tasting", while U3 in "picture card" expositions. These are two examples where ignoring composite term information (and including in the results irrelevant documents simply containing "tasting" or "picture") can seriously affect precision (24% and 12% for second baseline, compared to 90% and 92% of the semantic engine).

The results, especially for U3-U6 which contain a larger number of terms, also benefit from synonyms, related terms and class management. For instance, U3 contains terms such as "exposition" which are correctly matched with several documents about "exhibitions" that, otherwise, would have gone unnoticed. User U4 is typically interested in "clothing" items and in motor car racing: documents about Ferrari Formula One "shirts" and "sweaters" are correctly retrieved mainly because those terms are found related to "clothing" in WordNet, and the documents themselves have also been associated to such IPTC classes as "motor car racing" (similarly to some browsed documents). Users U5 and U6 previously browsed mainly art and food documents, respectively; identifying the synonymy/relatedness of such terms as "church" and "cathedral", "carving" and "sculpture" (U5) and "cheese" and "parmesan", "food store" and "food shop" (U6) guarantees very good recall/precision (over 90%), differently from the baselines.

Finally, we deepened the effectiveness analysis by considering the actual rankings of the retrieved documents. Being the number of potential suggestions very high, it is essential to evaluate whether the best suggestions are returned in the top positions (i.e. the weighting scheme is effective), especially for proactive cases. Figure 4 shows the normalized Spearman

footrule distance [6] between the retrieved and the ideal ranking for U5. As we can see, the curves of the semantic engine ("Std" in figure) are the lowest ones, meaning the least distance to the optimal ranking, while syntactic and non-weighted baselines are not effective in providing the best suggestions first. In particular, the fused ranking takes the best from the class and text rankings, which together seem able to well capture the user interests. For instance, taking U5 as a representative example, documents classified as "monument and heritage site" (IPTC) are deemed of interest to the art-focused user; text ranking also significantly contributes by promoting documents describing specific artistic sites such as a "clock tower", a term with an high weight in the user profile. Due to lack of space we do not show detailed analyses for the other users, however we found that the good performance of U5 is fully representative of the others.

5. Concluding remarks

Several works in the literature have highlighted the benefits of managing context information and/or proposed techniques and applications exploiting context-awareness capabilities ([2, 3, 4, 7]). In particular, a few works are directed towards context modeling, representation, and effective handling, aspects of particular interest to AMBIT. For instance, [3] proposes to design a context management system which is not application-dependent, [7] proposes an architectural framework for context data management, while [15] reports the result of a study on various context modeling and management approaches. However, most of the approaches in the literature primarily focus on specific aspects such as external user information, and/or do not consider the semantics of the context.

The techniques presented in this paper will serve as the foundations of the AMBIT Semantic Engine and are designed to be a first step in overcoming these limitations while being general enough to support different application scenarios. More specifically, they are focused on exploiting the textual information deeply characterizing both the documents to be retrieved and the user information, taking into account the users' "history", i.e. past navigated documents and requests. The proposed semantic similarity techniques leverage on the strengths of both classic information retrieval and of knowledge-based and classification techniques, adapted and extended from different contexts, from information disambiguation [9] to the querying of heterogeneous information in digital libraries and PDMSs [10] and assisted software engineering [11].

AMBIT has just started and the presented approach is one part of the final picture. Future work will include the design of the other components of the whole architecture, the possibility to "personalize" the retrieved information, for instance by highlighting the sections which should be the most interesting to the user, the study of additional aspects of a user profile and the extension of the semantic framework with additional relevant similarity techniques based on them. Finally, an actual real-world experimentation will be performed in the upcoming project evaluation phase in a number of case studies proposed by the project industrial partners in their field of expertise. This will help us in obtaining useful suggestions about the quality of the proposed techniques and their improvement, while, at the local level of our territory, the newly created context-dependent services is expected to help to improve the offer and to increase the volume of business of the partners.

References

[1] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. The Web never forgets: Persistent tracking mechanisms in the wild. In *21st ACM Conference on Computer and Communications Security*, Nov 2014.

[2] M. Baldauf, S. Dustdar, and F. Rosenberg. A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2:263–277, 2007.

[3] C. Bolchini, G. Orsi, E. Quintarelli, F. A. Schreiber, and L. Tanca. Context modeling and context awareness: steps forward in the context-addict project. *Bulletin of the Technical Committee on Data Engineering*, 34:47–54, 2011.

[4] G. Cabri, L. Leonardi, M. Mamei, and F. Zambonelli. Location-dependent Services for Mobile Users. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems And Humans*, 33(6):667–681, 11 2003.

[5] G. Cabri, M. Leoncini, and R. Martoglia. AMBIT: Towards an Architecture for the Development of Context-dependent Applications and Systems. In *3rd ICCASA Conference*, 2014.

[6] P. Diaconis and R. L. Graham. Spearman's footrule as a measure of disarray. *Royal Statistical Society Series B*, 32(24):262–268, 1977.

[7] P. Falcarin, M. Valla, J. Yu, C. A. Licciardi, C. Frà, and L. Lamorte. Context data management: An architectural framework for context-aware services. *Serv. Oriented Comput. Appl.*, 7(2):151–168, June 2013.

[8] C. Leacock and M. Chodorow. *Combining Local Context and WordNet Similarity for Word Sense Identification*, chapter 11, pages 265–283. The MIT Press, May 1998.

[9] F. Mandreoli and R. Martoglia. Knowledge-based sense disambiguation (almost) for all structures. *Information Systems (Information)*, 36(2):406–430, 2011.

[10] F. Mandreoli, R. Martoglia, W. Penzo, and S. Sassatelli. Data-sharing p2p networks with semantic approximation capabilities. *IEEE Internet Computing (IEEE)*, 13(5):60–70, 2009.

[11] R. Martoglia. Facilitate IT-Providing SMEs in Software Development: a Semantic Helper for Filtering and Searching Knowledge. In *SEKE*, pages 130–136, 2011.

[12] G. A. Miller. WordNet: A Lexical Database for English. *Communication of the ACM*, 38(11):39–41, 1995.

[13] M. E. Renda and U. Straccia. Web metasearch: Rank vs. score based rank aggregation methods. In *Proceedings of the 2003 ACM Symposium on Applied Computing*, pages 841–846, 2003.

[14] G. Salton and C. Buckley. Term-Weighting Approaches in Automatic Text Retrieval. *Inf. Process. Manage.*, 24(5):513–523, 1988.

[15] N. M. Villegas and H. A. Mller. Managing dynamic context to optimize smart interactions and services. In M. Chignell, J. Cordy, J. Ng, and Y. Yesha, editors, *The Smart Internet*, volume 6400 of *Lecture Notes in Computer Science*, pages 289–318. Springer Berlin Heidelberg, 2010.