A GENERAL APPROACH TO EXPLOIT ASPECTS OF INTELLIGENCE ON THE WEB

Laura Burzagli and Francesco Gabbanini
Institute of Applied Phisics, Italian National Research Council, Via Madonna del Piano 10, Florence, Italy

Keywords: Semantic web, Web 2.0, Natural language processing, Design for all.

Abstract: This contribution discusses the architecture of a software system that can be adopted to leverage the characteristics of Web 2.0 and Semantic Web, in order to make efficient usage of information. Key aspects on the implementation of a reusable framework are discussed, and the effectiveness of the approach is illustrated in an example scenario, in the context of inclusive e-Tourism.

1 INTRODUCTION

Intelligence is becoming a challenging and compelling functionality to cope with the evolution and the increased complexity of the Web, which is now an interactive ubiquitous information system that leverages the wisdom of many users and makes it possible to reuse data through mashups. From the perspective of users, this means having at the disposal a wealth of (poorly structured) information, which is increasingly provided by users themselves, and services that are useful in a variety of domains. In order to introduce intelligence in this environment, so as to fully exploit its potential and to make efficient usage of information, this contribution discusses the architecture of a software system that can be adopted to leverage the characteristics of two forms in which intelligence is generally recognized to manifest itself: Web 2.0 and Semantic Web. Key aspects on the implementation of a reusable framework to manage collective knowledge are discussed, and the effectiveness of the approach is illustrated in an example scenario, in the context of inclusive e-Tourism.

2 SW AND WEB 2.0

The ICT scientific community has started to study how the two different expressions of intelligence given by Web 2.0 and Semantic Web might come to a convergence (Heath and Motta, 2008; Yesilada and Harper, 2008; Ankolekar at al., 2008). This convergence leads to merge two different worlds. On the one side, the world of human participation and interaction between users, giving origin to the so called *collected intelligence*, that constitutes a peculiarity of Web 2.0. On the other side, the domain of *well-structured information*, and the capability of uncovering relations between concepts, which are generally recognized as strengths of the Semantic Web.

Whereas the massive amount of unstructured information provided by wide communities of Web 2.0 users benefits from being interlinked and structured using Semantic Web techniques, the Semantic Web would be of limited value if its ontologies were not populated with individuals and relations: for this reason it may take profit from the availability of large amount of data to be aggregated, provided by Web 2.0.

This idea has led to the construction of general approaches for the convergence of Web 2.0 and Semantic Web. In the following section, a software architecture of a *Collective Knowledge Management System* is described to exploit the convergence, building upon concepts presented in Burzagli et al., 2010.

3 DESCRIPTION OF THE ARCHITECTURE

The architecture of the system is a classical three-tier one, as outlined in figure 1, where the user interface layer is made up with structures allowing

Burzagli L. and Gabbanini F..
A GENERAL APPROACH TO EXPLOIT ASPECTS OF INTELLIGENCE ON THE WEB .
DOI: 10.5220/0003183104470450
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 447-450
ISBN: 978-989-8425-41-6

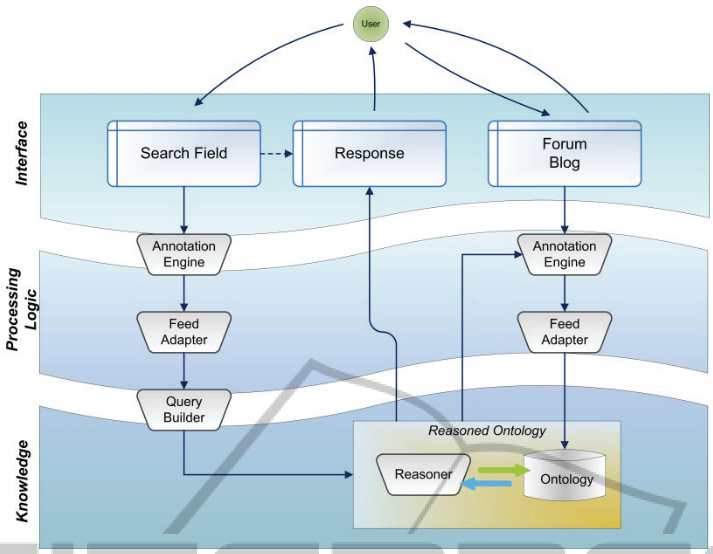

Figure 1: architecture of the Collective Knowledge Management System.

users to ask for information (either guided by forms or expressed in natural language) to the system and to read results. Moreover, it makes available social networking tools such as Forums, Blogs or Reviewing Systems that allow collecting user generated content.

A key layer of the architecture is represented by the Processing Logic layer, which includes an "Annotation Engine", and a "Feed Adapter", which operates as an annotation-ontology mapper. These two blocks, taken together, are capable of interpreting data that represent human intelligence (such as posts on social networking tools), using automatic learning techniques, and to insert these data in a hierarchical structure described by an ontology.

As a result, the third layer ("Knowledge Layer") is continuously and automatically augmented with the system's use, thanks to the discussions that take place between users. In this way the system is able to interact with the user in providing ever more personalized and pertinent information, thus optimizing the process of information fruition. This information is made available for browsing and searching by an ontology driven search engine that performs searches within a Reasoned Ontology.

Starting from the outlined architecture, fundamental structural blocks were identified in order to come to an implementation that has made use of available products, which were extended and integrated in order to form a reusable framework. This can be considered a first demonstration of the technological feasibility of the approach.

3.1 Implementation Details

This section discusses implementation details of a framework based on the general scheme described in section 3.

The reference platform on which the framework is based is the Java Platform, Standard Edition 6 (Java SE 6). Within the architecture, focus was put on the Processing Logic layer, where 2 distinct blocks can be identified:

- Language Processing block, in which contents originated by users are processed and annotated in order to provide coherent and structured inputs to the Knowledge layer (through the Feed Adapter), which uses them to generate new knowledge and enrich the ontology (potentially with new entities/properties and new individuals); implementation of this block is based on the General Architecture for Text Engineering (GATE, see Cunningham et al. (2002), Maynard et al. (2008)), which provides an object-oriented framework implemented in Java to embed language processing functionality in diverse applications.

- Feed Adapter, which was implemented from scratch as a middleware that provides a bridge from annotations produced by the Language Processing Block and the Knowledge Layer; the Feed Adapter provides interfaces that allow integration with a variety of frameworks for ontology manipulation, storage, inference and querying (e.g. the Sesame framework, is being used for the example described in section 4).

4 DESCRIPTION OF THE EXAMPLE

In order to better illustrate benefits coming from the use of a system combining the power of semantic intelligence and collective intelligence, a viable case study/application in the domain of inclusive tourism is being implemented using the framework described in the previous section.

The example focuses on inclusive e-Tourism and builds on an ontology whose core was built from scratch. Ontologies for inclusive tourism were developed within European projects such as ASK-IT (http://www.ask-it.org/) and Oasis (http://www.oasis-project.eu/). These ontologies often use categories to describe entities in terms of their accessibility (ex: wheelchair users/upper limb impaired users/lower limb impaired users). However, this kind of categorization was considered to be a bit rigid: for example it does not seem to be suitable to cope with preferences and requirements expressed by elderly people, which often have a mix of problems that characterize disabled people, even if in weaker forms. Therefore, work carried out in the CARE project (Città Accessibili delle Regioni Europee – Accessible Cities in the Regions of Europe) was used in a process of integration of an ontology in which we establish a set of relevant features that characterize indoor (and, partly, outdoor) environments so as to describe them in detail.

These characteristics are matched to the request of users to give a list of appropriate results, by means of a SPARQL capable search engine. This gives the possibility of performing very detailed and expressive searches.

The followed approach is thus in line with the Design for All approach because it is suitable to cope with preferences and requirements expressed by all people (including, for example, the elderly). Moreover, dropping categorization, avoids incurring in the eventuality that tourist resources that are classified as not being suitable to a user with definite characteristics are in fact suitable, because categories have been established in a too coarse way.

In other words, following a holistic approach by giving a detailed description of physical spaces may avoid incurring in misclassification of resources due to the fact that only a limited set of aspects are taken into account. This is also a field in which information contributed by the Web 2.0 may prove to be valuable.

4.1 Evaluating the Web 2.0 Contribution

In order to assess if the outlined approach helps improving performances in some specific tasks (which in this example consists in selecting a suitable accommodation), a correct evaluation process is being set up.

Part of the evaluation process overlaps with that aiming to evaluate Ontology Learning, and the scientific literature (see Buitelaar and Cimiano, 2008) contains pointers to papers dealing with the twofold aspects in which ontology learning evaluation consists of: evaluation of the ontology learning algorithm itself; task based evaluation, i. e. evaluation in the running application for which the ontology is engineered.

With reference to the first evaluation type, this is mainly technical, and a number of tests have been set up in order to tune up algorithms and to write grammars that are able to catch concepts that are relevant for the e-Tourism domain. As the framework uses at the moment JAPE based NLP techniques implemented in GATE, the results depend highly on how JAPE grammars are written. Availability of high quality grammars results in better ontology enrichment capabilities.

As for the second evaluation type, it is certainly the most important because it measures whether the approach actually brings improvements in the domain for which it was engineered, thus giving a measure of the success of the overall service. However, it is more difficult to cope with in a context such as the one we outlined in the previous section, because ontology learning and population are based on Web 2.0 corpora. This aspect adds a further degree of complexity in that corpora are continuously augmented and modified by users. In this case it would be useful to assess not only if a certain task is improved by using an ontology learning process, but also which is the actual added value that Web 2.0 brings. Without this, it would be difficult to assess the added value given by Web 2.0, in comparison with, for example, any other corpus collected by experts.

5 FUTURE DIRECTIONS

On the implementation side, work will regard refining JAPE based NLP techniques and enriching the NLP block by integrating different techniques into it (for example, starting from those described by Zablith et al., 2009). On the theoretical side, efforts will focus on how to cope with inconsistent assertions that the system may attempt to insert into the ontology. These are inevitably generated during ontology enrichment processes based on background knowledge coming from the web, and the topic is receiving attention by the scientific community (Sabou et al., 2009). As for the evaluation of the actual added value given by the Web 2.0 to ontology evolution, it is being investigated in the context of the so called "Task-based Approaches" for the evaluation of ontology learning (Dellschaft and Staab, 2008). It is to be noted that in general, the evaluation of Web 2.0 impact on the quality of information is an open problem, for which few references exist: a starting point could be the study of Giles (2005) on the comparison of accuracy between Wikipedia and Encyclopaedia Britannica. Following this model a comparison could be hypothesized between a service like the one we are proposing and those offered by one of the many booking platforms present on the market (like Booking.com or Expedia). Clearly, suitable metrics have to be set up and it must be observed that a correct evaluation should be conducted only after the service has been up and running for a certain time, in order for the corpus to reach a "critical mass".

6 CONCLUSIONS

The paper discusses the architecture of a software system that allows combining strengths of Web 2.0 and Semantic Web in order to make efficient usage of information. Some details on implementation aspects are discussed, and a possible application is illustrated in the context of inclusive e-Tourism. While the approach seems interesting, more work is certainly required to make it more mature: future directions regarding implementation and research issues are pointed out, of which the most challenging seem to be those regarding evaluation and handling inconsistencies.

REFERENCES

Ankolekar, A., Krotzsch, M., Tran, T., Vrandecic, D., 2008. The two cultures: Mashing up Web 2.0 and the Semantic Web. Web Semantics: Science, Services and Agents on the World Wide Web 6 (1), 70-75.

Buitelaar, P., Cimiano, P. (Eds.), 2008. Ontology Learning and Population: Bridging the Gap between Text and Knowledge. Vol. 167 of Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam.

Burzagli, L., Como, A., Gabbanini, F., 2010. Towards the convergence of Web 2.0 and Semantic Web for e-Inclusion. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (Eds.), Computers Helping People with Special Needs. Vol. 6180 of Lecture Notes in Computer Science. Springer, pp. 343-350.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., 2002. GATE: A framework and graphical development environment for robust NLP tools and applications. In: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics.

Dellschaft, K., Staab, S., 2008. Strategies for the evaluation of ontology learning. In: Buitelaar, P., Cimiano, P. (Eds.), Ontology Learning and Population: Bridging the Gap between Text and Knowledge. IOS Press, Amsterdam.

Giles, J., 2005. Internet encyclopaedias go head to head. Nature 438 (7070), 900-901.

Heath, T., Motta, E., 2008. Ease of interaction plus ease of integration: Combining Web 2.0 and the Semantic Web in a reviewing site. Web Semantics: Science, Services and Agents on the World Wide Web 6 (1), 76-83.

Maynard, D., Li, Y., Peters, W., 2008. NLP techniques for term extraction and ontology population. In: Proceedings of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge. IOS Press, Amsterdam, pp. 107-127.

Sabou, M., d'Aquin, M., Motta, E., 2008b. Exploring the semantic web as background knowledge for ontology matching. In: Spaccapietra, S., Pan, J., Thiran, P., Halpin, T., Staab, S., Svatek, V., Shvaiko, P., Roddick, J. (Eds.), Journal on Data Semantics XI. Vol. 5383 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, Berlin, Heidelberg, Ch. 6, pp. 156-190.

Yesilada, Y., Harper, S., 2008. Web 2.0 and the Semantic Web: hindrance or opportunity? W4A - international cross-disciplinary conference on web accessibility 2007. SIGACCESS Accessibility and Computing. (90), 19-31.

Zablith, F., Sabou, M., d'Aquin, M., Motta, E., 2009. Ontology evolution with evolva. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (Eds.), The Semantic Web: Research and Applications. Vol. 5554 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, Berlin, Heidelberg, Ch. 80, pp. 908-912.