Annotating Accommodation Advertisements using CERNO

Nadzeya Kiyavitskaya[a], Nicola Zeni[a], Luisa Mich[b],
James R. Cordy[c], and John Mylopoulos[d]

[a] Dept. of Information and Communication Technology, University of Trento, Italy
{nadzeya, nzeni}@dit.unitn.it; john.mylopoulos@unitn.it

[b] Dept. of Computer and Management Sciences, University of Trento, Italy
luisa.mich@unitn.it

[c] School of Computing, Queens University, Kingston, Canada
cordy@cs.queensu.ca

[d] Dept. of Computer Science, University of Toronto, Ontario, Canada
jm@cs.toronto.edu

Abstract

There has been great interest in applying Semantic Web technologies to the tourism sector ever since Tim Berners-Lee introduced his vision. Unfortunately, there is a major obstacle in realizing such applications: tourist (or other) information on the Web has to be semantically annotated, and this happens to be a very time- and resource-consuming process. In this work we present the application of a lightweight automated approach for the annotation of accommodation advertisements. The annotation tool, called Cerno, allows for annotation of text according to a predefined conceptual schema. Resulting annotations are stored in a database, allowing users to quickly find the best match to personal requirements. To evaluate our framework, we have conducted a series of experiments that support the efficacy of our proposal with respect to annotation quality and fulfilment of user information needs.

Keywords: semantic web, Cerno, semi-automatic annotation, accommodation ads

1 Introduction

The Semantic Web has been proposed as an extension of the World Wide Web where information is explicitly given well-defined meaning through semantic annotations. This vision has provoked great interest in the tourism sector since it was first proposed by Tim Berners-Lee in 1999 [1]. This interest is motivated largely by the importance of tourism in the world economy (for the U.S. see [2], also [3] for Europe). Moreover, in the last years, the percentage of people using the Internet as a medium to find a destination for vacations and reserve related services has been increasing steadily [4].

Potential applications based on the availability of semantically annotated information in tourism are described in [5], highlighting the feasibility of applying existing Semantic Web-based tools for next-generation tourism information systems. From a user viewpoint, the key idea is to allow the tourist to quickly retrieve all necessary information for organizing a vacation from the Web. This information relates to different phases of vacation planning - selection of a destination, finding the way to get there, choosing accommodation facilities, etc. - and must be provided in both efficient and effective ways. Thus far, users' experience in searching the data related to vacation planning shows that this process requires a long time and often gives unsatisfactory results. In this work we propose to demonstrate the merit of semantic web technologies as an instrument to improve the quality of service through a specific case study.

Many research projects work on implementing tools supporting the semantic annotation of documents, such as Kim [6] and SemTag [7], to name just two of the most known. These projects are characterised by different levels of automation of the annotation process. Available tools range from user-friendly interfaces facilitating hand annotation, to semi-automatic annotation tools where the user intervenes in one or more phases, to systems automatically annotating documents. For the applications in the tourism area the first approach – "computer supported" manual annotation – cannot be applicable, because of the large number of documents to be annotated and the high cost of keeping up with the rapid changes in their content. On the other hand, given the state-of-the-art in the systems for completely automatic semantic annotation, existing tools do not provide satisfactory results for arbitrary content, but only for relatively limited domains.Thus, the challenge is to develop annotation environments for complex sectors such as tourism that produce good quality results within reasonable time and resource limits. To address this trade-off we have to identify the requirements that the annotation tool has to satisfy for each specific task. In some cases the priority is to find all possible answers even if the tool provides in addition a few incorrect answers. For other tasks it is more important to select only correct matches. Using terminology from information retrieval, this means that in the first case *recall* is given greater importance, whereas in the second case we are interested in higher *precision* results. For example, for a tourist looking for an accommodation close to the Coliseum in Rome, it is more important to have at least one suitable answer. While for an official tourist operator providing information about lodging possibilities in Rome, it is crucial to have many accommodation options, organised along one or more dimensions. Here, both high recall and precision are desired.

In this work we analyse the results related to the application of semantic annotation to accommodation ads in a city of art. We performed several experiments using a semi-automatic annotation tool [8] (named *Cerno*) to annotate accommodation advertisements. Cerno is based on a lightweight text pattern-matching approach that takes advantage of a structural transformation system TXL [9]. The architecture and the performance of the tool are described in [10]. In this paper we assume the viewpoint of a web engineer that has to evaluate Cerno's ability to support the development of semantic annotation based applications.

From a linguistic viewpoint, online accommodation advertisements and similar documents pose a number of particular difficulties:

- partial and malformed sentences (for example, "30 square meters studio apt. in Rome center near FAO");
- abbreviations and short-forms ("Furn./equipp.");
- location-dependent vocabulary: names of geographical objects, both proper nouns ("Colosseum") and common nouns ("campo");
- presence of foreign language terms ("via", "Strasse", "Policlicnico");
- monetary units ("Euro 73.00/83.00/93.00 p.n.", "€2000");
- date and time conventions ("from the 15/20th of july 2006").

From the functional viewpoint, such advertisements are present in various kinds of websites publishing classified ads. These websites can belong to B2C (business to consumer) applications, where owners of a website offer space for publishing announcements, or to C2C (consumer to consumer) applications, where the users interact directly, for instance, within a virtual community. Both cases involve *non-intermediate* tourism (i.e. the segment of tourists who do not defer to on an intermediary, in other words, "do-it-yourself" tourists) where the Web is used in order to make visible and promote individual offers.

Searching these kinds of ads is very often a real test of patience for the user who has to look for the information in long lists of accommodation ads. Analogous problems can emerge when finding a hotel on tourist portals: most of them offer a keyword-based search; an example of more sophisticated search tool can be seen in the European tourism portal (www.visiteurope.com [Sept. 7, 2006]), which enables "semantic queries" (for example, find all accommodation in the vicinity of the Coliseum). However, solutions of this type demand large investments and are affordable only for large organisations. Taking into account both linguistic and functional issues, it is important to explore lightweight tools for semantic annotation.

The remainder of this paper is structured as follows: Section 2 describes the features of Cerno and explains how it addresses the problems of the domain of accommodation ads. Section 3 presents the details of applying Cerno to annotate accommodation ads from different websites. Section 4 contains a detailed analysis of the annotation results with respect to the information needs of tourists. Section 5 reviews related work in the field, and finally, the concluding section provides a list of open issues and future work.

2 Semantic Annotation Using Cerno

Cerno is based on the parsing and structural transformation system TXL [9], which allows by-example rapid prototyping of language descriptions, tools and applications. The architecture of Cerno (Fig. 1) is based on the LS/2000 software analysis architecture [14], generalized to allow for easy parameterization by a range of semantic domains. To annotate input documents, Cerno uses context-free grammars, generates a parse tree, and applies transformation rules to generate output in a target format [10]. The architecture of Cerno factors out reusable domain independent knowledge such as the structure of basic entities (email and web addresses, dates, and other word-equivalent objects) and language structures (document, paragraph, sentence), while allowing for easy change of semantic domain, characterized by vocabulary (category word and phrase lists) and semantic model (entity-relationship or UML class diagram and interpretation).

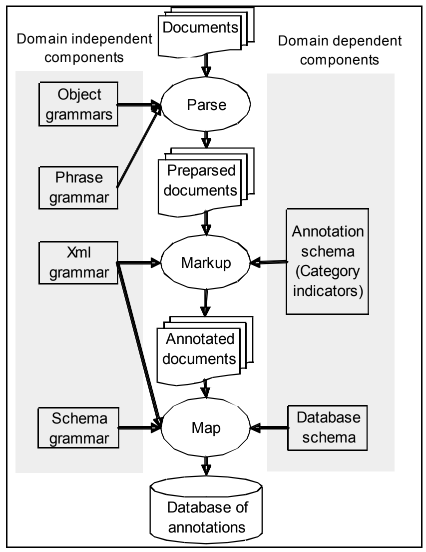

Fig. 1. Architecture of our semantic annotation process.

The annotation process consists of three main phases. In the first phase, a context-free grammar is used to efficiently obtain an approximate phrase structure parse of the source text. This stage results in a deterministic maximal parse. As part of this first stage, basic entities are recognized. The parse is linear in the length of the input and runs at compiler speeds [10].

In the second phase, initial semantic annotation of the document is derived using a wordlist file specifying both positive and negative indicators for semantic categories. Indicators can be both literal words and phrases and names of parsed entities. Phrases are marked up once for each category they match – thus at this stage a sentence or phrase may end up with many different semantic annotations. Vocabulary lists are derived from the semantic model for the target domain. This stage uses the structural pattern matching and source transformation capabilities of the TXL engine similarly as for software markup to yield a preliminary marked-up text in XML form.

The third stage uses the XML marked-up text to populate an XML database schema, derived from the semantic model for the target domain. Sentences and phrases with multiple annotations are copied, one for each different markup, before populating the database. In this way we do not prejudice one interpretation as being preferred. The final outputs are both the XML marked-up text and the populated database.

3 Semantic Annotation of Accommodation Ads

In our first experiment we used a set of several hundred advertisements for accommodation in Rome drawn from an online newspaper [11], see an example in Fig. 2.

```
Very elegant apartment located in Piazza Dante, just a walk from
Fosse Ardeatine and 10 minutes to Colosseum by bus (Bus stop in the
square. 75 smq in a charming, and full furnished environment. 1.200
euro a month, utilities not included. Write to
pseudonym@somewhere.it or phone to 123.1234567
```

Fig. 2. Sample accommodation advertisement.

To annotate these ads, we designed a conceptual schema which represents the information needs of a tourist (Fig. 3). Then, as the first step to produce the annotation schema, we derived the most general concepts of the model: *Accommodation Type, Owner, Contact, Facility, Term* (of availability), *Location,* and *Price*.

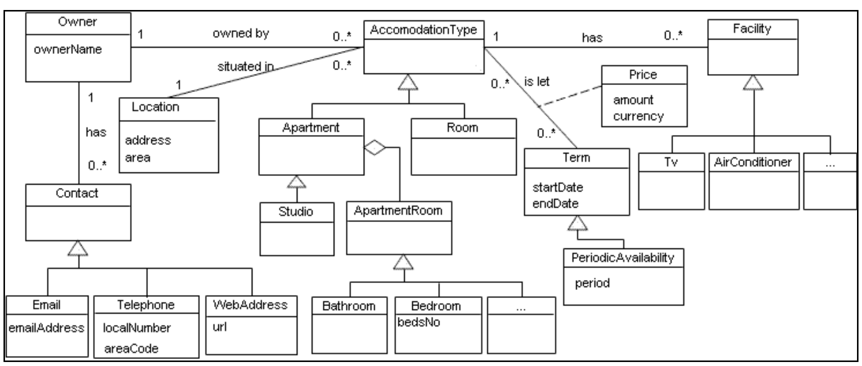

Fig. 3. Conceptual schema for accommodation ads represented in UML notation (http://www.uml.org/ [Sept. 7, 2006]).

First analysis of this list allows us to omit the concept of *Owner,* because this information is rarely provided in ads. Hence, the annotation schema derived consisted of the following concepts: *Type, Contact, Facility, Term*, *Location,* and *Price*. This schema is rather standard, given the nature of the documents. The goal of this experimental study was to identify and annotate the concepts in the advertisements according to the given conceptual schema, which was translated into XML (www.w3.org [Sept. 7, 2006]) for input to the Cerno tool. The tags, i.e. concepts, used for annotation are given in Fig. 4.

```
<ad>
      <location></location>
      <price></price>
      <contact></contact>
      <facility></facility>
      <term></term>
      <type></type>
</ad>
```

Fig. 4. Database template schema for accommodation advertisements.

To adapt our semantic annotation methodology to this experiment, the domain-related wordlists were constructed by hand from a set of examples. In order to make a realistic test of the "generality" of the method we restricted ourselves to some constraints. In particular, we avoided all proper names and locality-dependent words, and we did not pre-process the text of accommodation descriptions by formatting them or correcting errors. The output of the annotation was the annotated original advertisements (Fig. 5) and a database with one instance of the schema for each advertisement in the input.

```
<type><location>Very elegant apartment located in Piazza Dante, just
a walk from Fosse Ardeatine and 10 minutes to Colosseum by bus (Bus
stop in the square)</location></type>. <facility> 75 smq in a
charming, and full furnished environment </facility>.
<facility><price>1.200 euro a month, utilities not included
</price></facility>. <contact> Write to pseudonym@somewhere.it or
phone to 123.1234567 </contact>
```

Fig. 5. Example result of an annotated accommodation advertisement.

Next we performed a series of experiments in order to estimate the flexibility of the annotation tool to different geographic regions. In particular, we analysed accommodation ads in Venice [12] and Paris [13]. Both of these posed new problems compared to the ads in Rome, for example different document structure and vocabulary, and missing information, such as contact and price of accommodation.

4 Analysis of the Results

4.1 Evaluation Measures

The performance of semantic annotation tools is usually evaluated similarly to information extraction systems, i.e. by comparing with a reference annotation and calculating recall, precision and other quality measures. In order to evaluate our experimental results, we calculated a number of measures for the tool's automated annotation compared to manually-generated annotations:

- *Recall* shows how well the tool performs in finding relevant items were found (i.e. the number of relevant items detected divided by the number of all relevant items in the collection):

$$recall = \frac{TP}{TP + FN} \quad (1)$$

- *Precision* shows how well the tool performs in not returning irrelevant items (i.e. the number of relevant items detected divided by the number of all items detected):

$$precision = \frac{TP}{TP + FP} \quad (2)$$

- *Fallout* measures how quickly precision drops as recall is increased (i.e. the number of irrelevant items detected divided by the number of irrelevant items in the collection):

$$fallout = \frac{FP}{FP + TN} \quad (3)$$

- *Accuracy* measures how well the tool identifies relevant items and rejects irrelevant ones (i.e. the number of correctly detected and correctly rejected items divided the number of all items);

$$accuracy = \frac{TP + TN}{N} \quad (4)$$

- *Error* rate demonstrates how much the tool is prone to accept irrelevant items and reject relevant ones (i.e. the number of incorrectly detected and incorrectly rejected items divided the number of all items):

$$error = \frac{FP + FN}{N} \quad (5)$$

In these formulas N is the total number of test items $N = TP + FP + FN + TN$, where:

- TP (true positives) is the number of items correctly assigned to the category;
- FP (false positives) is the number of items incorrectly assigned to the category;
- FN (false negatives) is the number of items incorrectly rejected from the category;
- TN (true negatives) is the number of items correctly rejected from the category.

All these measures were used during the analysis of the results.

4.2 Results of the Annotation

In order to estimate the quality of automatic annotations in our experiment, Cerno and a human marker were given one hundred advertisements to annotate. These documents were different from the training set used to tune the tool for the domain (a second human annotation was used to evaluate the confidence level of the reference annotation: differences were lower than 1% for all the measures). The tool was then compared against the human annotation to calculate the five measures described in this section (Table 1).

Table 1. Evaluating tool annotation (Rome, 100 ads)

Concept / Measure	Contact	Facility	Location	Price	Term	Type	*Average*
Recall	99.07	96.69	77.00	97.09	76.19	93.97	*90.00*
Precision	100.00	96.15	98.72	96.15	91.43	96.46	*96.49*
Fallout	0.00	2.62	0.29	1.16	0.74	1.20	*1.00*
Accuracy	99.78	97.10	94.64	98.44	97.10	97.54	*97.43*
Error	0.22	2.90	5.36	1.56	2.90	2.46	*2.57*

From the annotation tool viewpoint, these results are very promising. In fact, without local knowledge and using a very small vocabulary and only few TXL rules for non-trivial semantic concepts (most of which are not covered by traditional information extraction systems) we obtained results comparable to some of the best heavyweight annotation methods, albeit on a limited domain. The tool was also very fast, handling 100 advertisements in about 1 second on a 1 GHz PC.

Next we measured the effect of the initial automated annotation of the tool on human annotation productivity. The time taken by an unassisted human marker to semantically annotate a new sample of 100 advertisements was measured, and compared to the time taken by the same human marker when asked to correct the automated markup created by the tool. In this first evaluation the human annotator was observed to use 78% less time to mark up text with assistance than without, a significant saving. Because the system was shown in the first evaluation to be more aggressive than humans in markup, the majority of the correction work was removing markup inserted by the tool.

From an application viewpoint, the results in Table 1 represent a useful input to the designer of a semantic annotation based application. In fact, the annotation tool exhibited in average high level of recall, precision and accuracy; and a quite low level of fallout and error (the last column of Table 1). However, the results are not homogeneous for the different concepts. According to the table, the information on *contact* and *price* was more accurate than the information on the other concepts.

However, to interpret these results we should take into account that the relevance of these entities depends on tourist needs. To this end we adapted existing data to the information needs of a tourist looking for any kind of accommodation obtained from a large survey [15]. According to their list of the “relevant information categories” for the accommodation seeking problem, the subset of concepts used for our annotations, can be classified as follows, starting from the most important: *Location, Price, Type, Facilities*.

Contact and *term* were not considered in the study, but we can assume that *contact* is more relevant than *term* (availability can be checked only if the contact is correctly given).

In this context, in respect to that classification, we could say that for a tourist looking for accommodation, the annotations of Cerno satisfy the requirements for the most relevant concepts, *contact* and *location* (with a high precision, even if the recall is not 100% we are sure that we would obtained only suitable accommodations, even if we will not find all of them). The results for the other concepts are also adequate (accuracy ranges from 94.64 to 99.78) for a single tourist search. However, for an exhaustive search, for example to classify all the accommodation in a given area of Rome for a tourist guide, we have to take into account that the system was not able to identify more than 23% of the location related information in the ads (given that ads could contain more than one location, this result represents an upper limit for the missing ads), so that the tool annotation would have to be revised by an human annotator.

The (small) differences in the results are mainly due to the level of generality of the concepts: for example, for *facility,* on one hand there are a number of facilities that could be found in an accommodation; on the other hand, there are no linguistic forms (as for *contact*) or specific symbols (as for *price*); and for *term*, the results depend a lot on how this notion is defined (for example, phrases like "Studio available for holidays" and "Reductions for long term stay" implicitly contain information about the time for rent, but human markers may ignore it as not being relevant enough).

Results obtained for the accommodation ads in Paris and Venice are given in table 2 and 3, respectively.

Table 2. Evaluating tool annotation (Paris, 10 ads)

Concept / **Measure**	Contact	Facility	Location	Price	Term	Type
Recall	-	78.79	69.57	50.00	100.00	81.25
Precision	-	100.00	88.89	50.00	66.67	100.00
Fallout	-	0.00	5.00	1.64	1.61	0.00
Accuracy	-	88.89	85.71	96.83	98.44	95.24
Error	-	11.11	14.29	3.17	1.56	4.76

Table 3. Evaluating tool annotation (Venice, 10 ads)

Concept / **Measure**	Contact	Facility	Location	Price	Term	Type
Recall	-	92.86	58.33	100.00	100.00	100.00
Precision	-	92.86	100.00	90.91	50.00	100.00
Fallout	-	2.63	0.00	2.38	1.96	0.00
Accuracy	-	96.15	80.77	98.08	98.08	100.00
Error	-	3.85	19.23	1.92	1.92	0.00

The results show excellent quality rates for some concepts while there is a drop in performance for others. While some of the differences are due to the different styles of the ads (e.g., they were published on a different website for Paris), most of them are related to the geographical and cultural differences described in the introduction. Thus, these data give useful hints on how to tune the tool's domain-dependent components to obtain better quality results.

5 Related Work

There are many other systems and projects related to semantic annotation; here we have room to name only two of the best known, to set our work in context. SemTag [7] annotates large numbers of web pages with concepts from TAP ontology, using corpus statistics to improve the quality of tags. The TAP knowledge base contains lexical and taxonomic information about such objects as musicians, movies, authors, athletes, states and cities, and other generic terms. SemTag detects the occurrence of these entities in web pages and disambiguates them. The KIM platform [6] is an application for automatic ontology-based named entities annotation, indexing and retrieval. In KIM, as well as in SemTag, semantic annotation is considered as the process of assigning to the entities in the test links to their semantic descriptions, provided by ontology. The platform is based on GATE (General Architecture for Text Engineering). The main contribution of KIM is recognition of named entities with respect to the KIMO ontology, which covers such items as *Organisation, Person, Location, Event, TimeInterval* and others. Another tool that has been used on a large-scale is SCORE [20], which integrates several information extraction methods, including probabilistic, learning, and knowledge-based techniques, then combines the results from the different classifiers.

In contrast to all these tools our method uses a lightweight robust context-free parse in place of linguistic analysis. Moreover, in Cerno there is no learning phase, instead the tool has to be tuned manually when being ported to a particular application. Cerno does not rely on a knowledge base of known named entities, rather it detects them in the style of software analyzers – basing on the structural and vocabulary context of these entities [8]. This advantage helps make our tool faster and less dependent on the additional knowledge sources.

Focusing on the tourism sector, in recent years, many large projects have been undertaken, especially in the area of ontology development. Among these is Harmonise [16], which aims at creating a comprehensive ontology for Tourism. Another group developing an ontology in this sector is DERI [17]. The DERI website gives also a long list of existing ontologies related to the tourism sector (University of Innsbruck: e-Tourism Portal: Ontology Collection [18]). One of the more recent initiatives is the Open Travel Alliance (http://www.opentravel.org/ [Sept. 7, 2006]), which works on producing XML message specification schemas for trading partners. Our project does not use any specific ontology, but rather uses a small conceptual schema which could be derived from or integrated with any of these more general ontologies, translating to a standard ontology language such as OWL (http://www.w3.org/2004/OWL/ [Oct. 24, 2006]).

Other projects have the goal to create an interactive environment which exploits semantically annotated documents to integrate the information available from different sources. For example, one of the few applications in tourism sector takes advantage of semantic annotations to enable dynamic packaging [19].

6 Conclusions

In this paper we have presented the application of Cerno, a lightweight tool for semantic annotation of web documents, to the annotation of accommodation ads. The results of the experiments, albeit preliminary, suggest that the quality of the output is adequate to develop web based applications supporting accommodation search. The main advantages using Cerno are the following:

- Cerno requires a limited use of external resources (no gazetteers, vocabularies, ontologies); if necessary, and for specific applications, the tool architecture allows the integration of such components with minimal effort.
- Cerno requires only limited computational resources, so that it could easily be adapted to "lightweight" interfaces to access tourist information or to online real-time applications.
- The effort to adapt Cerno to a new domain is minimal compared to what is required for other tools.

In addition, the comparison with data on the tourist information allows us to adopt a engineering approach that focuses on the feasibility of applications based on semantically annotated accommodation ads. For the designer of a semantic annotation based application the results of the experiments represent useful data suggesting how to adjust the tool to obtain adequate quality results for a particular task. Cerno has also been used to annotate tourist board websites with a very rich conceptual schema representing the knowledge of marketing experts with the goal of evaluating the communicative effectiveness of those websites.

Potential applications of the system presented in this paper are related to (a) the implementation of an interface to support accommodation search based on semantic annotation of documents (e.g., a web service usable both by human and artificial agents); (b) the integration of accommodations ads for a given city published on differently structured web sites. As regards further experimentation, the following topics need to be investigated: performance of the tool in case if external resources are used; identification of thresholds of the evaluation measures' values for a given application (for example, if the tool does not give "all" the accommodation ads that satisfy the tourist's needs, it is not a problem for the tourist, but it could be a problem for the owners of the missed accommodation).

References

1. Berners-Lee, T., Hendler, J., & Lassila, O. (2001) The Semantic Web: A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities. *Scientific American*, 284 (5), May 2001, 28-37.
2. Travel Industry Association of America (2004) Economic Research: Economic Impact of Travel and Tourism. December, 2004, URL: http://www.tia.org/Travel/econimpact.asp. Retrieved: Sept. 7, 2006.
3. European Commission. Tourism. Facts and figures about European tourism. URL: http://ec.europa.eu/enterprise/services/tourismeu.htm. Retrieved: Sept. 7, 2006.
4. European Travel Commission (2006) World and Regional Overview. 4 September 2006, URL: http://www.etcnewmedia.com/review/default.asp?SectionID=10. Retrieved: Sept. 13, 2006.
5. Maedche, A. & Staab S. (2002) Applying Semantic Web Technologies for Tourism Information Systems. In: K. Wöber, A. Frew, M. Hitz (eds.), Proc. of the *9th Int. Conf. for Information and Communication Technologies in Tourism (ENTER 2002)*, Innsbruck, Austria, 2002.
6. Kiryakov, A., Popov, B., Terziev, I., Manov, D., & Ognyanoff, D. (2005) Semantic Annotation, Indexing, and Retrieval. *Journal of Web Semantics*, 2005, 2(1), 49-79
7. Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., McCurley, K.S., Rajagopalan, S., Tomkins, A., Tomlin, J.A. & Zien, J.Y. (2003) A Case for Automated Large-Scale Semantic Annotation. *Journal of Web Semantics*, 2003, 1(1) 115-132
8. Kiyavitskaya, N., Zeni, N., Mich, L., Cordy, J.R. & Mylopoulos, J. (2006) Text Mining through Semi Automatic Semantic Annotation. To be published in Proc. of PAKM'2006.
9. Cordy, J.R. (2006) The TXL Source Transformation Language. *Science of Computer Programming*, 61 (3), August 2006, 190-210.
10. Kiyavitskaya, N., Zeni, N., Cordy, J.R., Mich, L., & Mylopoulos, J. (2005) Applying Software Analysis Technology to Lightweight Semantic Markup of Document Text. In Proc. *of Int. Conf. on Advances in Pattern Recognition (ICAPR 2005)*, Bath, UK, 2005, 590-600.
11. Live In Rome & Italy, Rome Ads - Classifieds for Rome, Italy URL: http://www.liveinrome.com/classifieds.htm. Retrieved: January 2005.
12. Live In Rome & Italy, Apartments in Venice. URL: http://www.liveinrome.com/apartments/venice.htm. Retrieved: March 2005.
13. Rentcom.info Property Portal. Property in France-Paris Apartments. URL: http://www.rentcom.info/paris-apartments/paris-apartments-2.html. Retrieved: March 2005.
14. Dean, T., Cordy, J., Schneider, K., & Malton, A. (2001) Experience using design recovery techniques to transform legacy systems. In Proc. of the *17 Int. Conf. on Software Maintenance*, 2001, 622-631
15. Hepp, M., Siorpaes, K., & Bachlechner, D. (2006) Towards the Semantic Web in E-Tourism: Can Annotation Do the Trick? In Proc. of the 14th European Conf. on Information System (ECIS 2006), June 12–14, 2006, Gothenburg, Sweden.
16. Dell'Erba, M., Fodor, O., Ricci, F., & Werthner H. (2002) Harmonise: A Solution for Data Interoperability. *In Proc. of the IFIP Conference on Towards The Knowledge Society: E-Commerce, E-Business, E-Government*, October 2002.
17. Digital Enterprise Research Institute (DERI). E-Tourism Working Group . URL: http://e-tourism.deri.at/. Retrieved: Sept. 13, 2006.
18. E-Tourism Working Group (2004) Ontology Collection in view of an E-Tourism Portal. 5 October 2004. URL: http://138.232.65.141/deri_at/research/projects/e-tourism/2004/d10/v0.2/20041005/ Retrieved: Sept. 13, 2006.
19. Cardoso, J. (2006) E-Tourism: Creating Dynamic Packages using Semantic Web Processes. W3C Workshop on Frameworks for Semantics in Web Services (http://www.w3.org/2005/01/ws-swsf-cfp.html), June 9-10, 2005, Innsbruck, Austria.
20. Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., & Warke, Y. (2002) Managing Semantic Content for the Web. *IEEE Internet Computing*, 2002, 6 (4) 80-87.