Benchmarking Usability of Annotation Tools for the Semantic Web

Diana Maynard

October 6, 2005

1   Introduction

This paper discusses and explores the main issues for evaluating annotation tools in terms of usability. Semantic annotation and ontology-based information extraction technologies form the cornerstone of text mining and many other applications for the Semantic Web. There has been a great deal of work in the last decade on evaluating systems in terms of performance, but there are many other aspects of such tools which also need to be evaluated, in particular in industrial settings where companies wish to compare different systems to know which one is best suited to their needs. In tne context of the Semantic Web, there is a huge range of possible use cases and scenarios for annotation tools, and requirements may be widely different for different users. Comparing or establishing the usefulness of different systems goes beyond the typical requirements for a traditional IE system in terms of performance, considering criteria such as functionality, scalability, accessibility and interoperability [Maynard, 2005]. In this paper, we describe some of the main requirements for annotation tools in terms of benchmarking usability.

2   Expected functionality

First, we briefly explain the expected functionality of ontology-based annotation tools, in terms of the minimum that they should be expected to achieve. Annotation tools may fall into several types: manual, semi-automatic or automatic. All three types follow the same basic principles and can be evaluated in a similar way, though various considerations need to be taken into account according to the type. The relative speed of annotation for different systems (both manual and automatic) is, for example, an important criterion. The objective of the tool is that given a corpus of text and an existing ontology, it should be able to create semantic metadata by populating the texts with instances from the ontology. In some cases it may also modify the existing ontology structure or content, for example by adding new concepts or reassigning instances, but this is not obligatory.

3 Interoperability

One very important requirement of annotation tools is that they should be interoperable, i.e. that they can be combined with other tools, systems, and datasets and used by different people with different requirements. In particular, the format of the results produced should be usable in other systems and applications, because the results of semantic annotation are not generally useful as a final product, but only when combined with other tools or systems such as information retrieval and other more specialised search facilities, question answering, data evaluation, technology watch and market monitoring, and so on.

Interoperability covers issues such as standoff vs inline annotation as the representation for the results produced, the format of the ontology (e.g. RDF, OWL, DAML+OIL etc), the platform and browser on which the tool runs, etc. A more detailed study of interoperability amongst annotation systems can be found in [Siberski, 2005].

4 General Usability

Usability is a criterion which is not generally awarded much importance in the research community, but which bears far more significance when evaluating tools for use in industry. General usability includes criteria such as ease of installation and quality of installation documentation, general quality, format and ease of access of documentation for running the software, ease of setup, general aesthetics, and simplicity/complexity of the tasks possible.

5 Accessibility

Software accessibility is essentially about making tools that are usable, easy to use and attractive to use for everyone (not just for people with disabilities). Generally, however, designing software with certain disabilities in mind covers the majority of cases for people with and without disabilities. Some of the most important examples of accessibility problems stem from inflexibility. A well designed tool will have options to change the user's preferences regarding colours, layout, font sizes and styles, and so on, and the ability to save and restore latest sessions, etc. Mouse alternatives are another important aspect, especially for people with RSI, motor and sight problems.

6 Scalability

The question of scalability generally reflects a tradeoff between the level of automation, the size of the corpus, and the quality of the final output. Systems which perform well and on large documents are unlikely to be fully automatic; systems which are fully automatic may be able to handle large documents but with lower performance. Other scalability issues concern storage and manipulation of large ontologies and knowledge bases, and processing speed when dealing with large volumes of data.

7 Reusability

Ideally, annotation systems should be reusable in a wide variety of contexts, i.e. they should work on different kinds of domains and genres. Semi-automatic systems which rely on some degree of manual annotation and/or training can usually be adapted to new domains and ontologies, but will need retraining by the user. Automatic methods, on the other hand, can represent many different views, and they change according to the ontology in question. Again there is usually a tradeoff between the two. Reusability is also linked with interoperability - the more interoperable the tool is, the more reusable it tends to be, because some or all of its components can easily be integrated into other systems.

8 Implementation

The criteria for benchmarking usability described in this paper have been put into practice in two ways within the KnowledgeWeb Network of Excellence. First, a benchmark suite of tools for evaluation is being developed, and reports and experiments are underway to measure the usability and performance of existing tools and applications for the semantic web, so that users have a means of comparing the products in which they are interested. Second, we have applied our own criteria to tools for the semantic web being developed by project partners, such as ASPL (Advanced Semantic Platform for Learning) [Dzbor and Stutt, 2005]. In particular we have created questionnaires, feedback forms and user testing experiments to ensure that these criteria are fulfilled. Further details and results of these will be forthcoming in the full paper.

9 Conclusions

In this paper we outline some of the most important usability features required by annotation tools for the semantic web, and show how these can incorporated in a benchmarking suite for evaluation. In particular, this helps industrial (and also academic) users to decide on the best tools for their needs. In the full paper we shall give more details and elaborate on some experiments we have carried out.

References

[Dzbor and Stutt, 2005] Dzbor, M. and A. Stutt, 2005. Prototype of advanced learning platform (ASPL-v1). Technical Report D3.3.3, KnowledgeWeb Deliverable.

[Maynard, 2005] Maynard, D., 2005. Benchmarking ontology-based annotation tools for the semantic web. In *UK e-Science Programme All Hands Meeting (AHM2005) Workshop "Text Mining, e-Research and Grid-enabled Language Technology"*. Nottingham, UK.

[Siberski, 2005] Siberski, W., 2005. Semantic web framework requirements analysis. Technical Report D1.2.2, KnowledgeWeb Deliverable.