

The Surplus Value of Semantic Annotations

Maarten Marx
ISLA, Informatics Institute, University of Amsterdam
The Netherlands
maartenmarx@uva.nl

ABSTRACT

We compare the costs of semantic annotation of textual documents to its benefits for information processing tasks. Semantic annotation can improve the performance of retrieval tasks and facilitates an improved search experience through faceted search, focused retrieval, better document summaries, and result grouping. Applications which summarize large collections of text or explain real world phenomena based on textual evidence may receive even more benefit from semantic annotations.

Semantic annotation creates surplus value if the annotated data can be used beyond any foreseen application. In particular by third parties linking your data by means of your semantic markup to other data with similar markup. We present a list of properties of the annotated data which optimize this surplus value. They are derived from the principle which states that annotation should facilitate the reuse of data in a mashup without information being lost or distorted.

For the Dutch House of Parliament we annotated the parliamentary proceedings based on this principle. Concrete examples from this data collection will illustrate the surplus value enhancing properties.

General Terms

Design

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous