UNIVERSITÀ DI PISA

Scuola di Dottorato in Ingegneria "Leonardo da Vinci"

Corso di Dottorato di Ricerca in
INGEGNERIA DELL'INFORMAZIONE

Tesi di Dottorato di Ricerca

Collaborative editing of knowledge resources for cross-lingual text mining

Autore:

Francesco Ronzano _____

Relatori:

Ing. Alessio Bechini _____

Dott. Andrea Marchetti _____

Anno 2011

ABSTRACT

The need to smoothly deal with textual documents expressed in different languages is increasingly becoming a relevant issue in modern text mining environments. Recently the research on this field has been considerably fostered by the necessity for Web users to easily search and browse the growing amount of heterogeneous multilingual contents available on-line as well as by the related spread of the Semantic Web. A common approach to cross-lingual text mining relies on the exploitation of sets of *properly structured multilingual knowledge resources*. The involvement of huge communities of users spread over different locations represents a valuable aid to create, enrich, and refine these knowledge resources. *Collaborative editing Web environments* are usually exploited to this purpose.

This thesis analyzes the features of several knowledge editing tools, both semantic wikis and ontology editors, and discusses the main challenges related to the design and development of this kind of tools. Subsequently, it presents the design, implementation, and evaluation of the Wikyoto Knowledge Editor, called also Wikyoto. Wikyoto is the collaborative editing Web environment that enables Web users lacking any knowledge engineering background to edit the multilingual network of knowledge resources exploited by KYOTO, a cross-lingual text mining system developed in the context of the KYOTO European Project.

To experiment real benefits from social editing of knowledge resources, it is important to provide common Web users with simplified and intuitive interfaces and interaction patterns. Users need to be motivated and properly driven so as to supply information useful for cross-lingual text mining. In addition, the management and coordination of their concurrent editing actions involve relevant technical issues.

In the design of Wikyoto, all these requirements have been considered together with the structure and the set of knowledge resources exploited by KYOTO. Wikyoto aims at enabling common Web users to formalize cross-lingual knowledge by exploiting *simplified language-driven interactions*. At the same time, Wikyoto generates the set of complex knowledge structures needed by computers to mine information from textual contents. The learning curve of Wikyoto has been kept as shallow as possible by hiding the complexity of the knowledge structures to the users. This goal has been pursued by both enhancing the simplicity and interactivity of knowledge editing patterns and by using natural language interviews to carry out the most complex knowledge editing tasks. In this context, TMEKO, a methodology useful to support users to easily formalize cross-lingual information by natural language interviews has been defined. The collaborative creation of knowledge resources has been evaluated in Wikyoto.

INDEX