Unifying Web-scale Search and Reasoning
from the Viewpoint of Granularity

Yi Zeng, Yan Wang
(International WIC Institute, Beijing University of Technology)

Abstract

Considering the time constraints, it is impossible to achieve absolutely complete reasoning results based on Web-scale data. Plus, for the same query, the same results may not meet the diversity of user needs since their expectations may differ a lot. One of the major solutions for this problem is to unify search and reasoning. From the perspective of granularity, this talk provides various strategies of unifying search and reasoning for effective problem solving on the Web. We bring the strategies of multilevel, multiperspective, starting point (inspired by the basic level advantage in Cognitive Science) from human problem solving to Web scale reasoning to satisfy a wide variety of user needs and to remove the scalability barriers. Concrete methods such as network statistics based data selection and ontology supervised hierarchical reasoning are applied to these strategies.

For the diversity issue: The strategy of starting point focuses on user specific background and the unification process is familiarity driven or novelty driven, and is obviously user oriented. Multilevel completeness strategy is with anytime behavior, and provides predictions of completeness for user judges when the user interacts with the system. Multilevel specificity strategy emphasizes on reasoning with multiple levels of specificity and users can choose whether to go into more specific or more general levels. Multiperspective strategy attempts to meet various user needs from multiple perspectives. For the scalability issue: In the multilevel completeness strategy, although the partial results may have low completeness, more important results have been searched out and ranked to the top ones for reasoning. In other words, more important results are provided as a possible way to solve the scalability problems. The starting point strategy also provides two methods to select important nodes for reasoning. The multilevel specificity strategy concentrates on the appropriate levels of specificity controlled by the knowledge hierarchy and does not get into unnecessary levels of data. Hence, under limited time, the reasoning task and time is reduced. The experimental results based on the SwetoDBLP dataset shows that the proposed strategies are potentially effective.

Acknowledgement