Exploring the Web for Real World Activities Related to User's Query

The web is full of information concerning activities. An activity is something a person can accomplish or something one can create. This information is concentrated on websites called How-To or DIY websites, where you can find articles detailing methods to accomplish any action a person can imagine, from the simplest to the really complicated ones. People can search for these activities in the traditional way using a search engine.

What our research focuses on is how we can allow a user interested in a general query to search for activities to accomplish related to that query. For example a user searching for the query coffee can find among others the activity "Wake up and exercise in the morning". It is an activity that does not directly contain the search query but is conceptually related to it.

We accomplish this by creating a system that performs Query Expansion on the search query, employing a knowledge base created by MIT called ConceptNet. The query is expanded and we obtain related words that are tagged depending on the relation to the original query and with every relation getting a different weight parameter. Then the How-To websites are searched with these related words and the resulting activities are returned to the user.

The order that the activities are presented in is important and for that we have employed optimization using genetic algorithms. We conducted a small experiment where we gathered data for a large set of queries and constructed the so called gold standard list for each query, which is the perfect order of activities for that query. The genetic algorithms optimize the relation weight parameters so that the final order of the resulting activities is as relevant as possible to the original by attempting to converge to the gold standard.

Finally we conducted a small evaluation comparing our optimized system with its un-optimized equivalent and simple web search of the How-To websites and it was concluded that in both cases the optimized system was superior to the other two. For the future we are planning to tweak the genetic algorithms implementation and conduct more experiments and finally perform a bigger evaluation of the system in its entirety.

References:

[1] Karapetsas E. Fukazawa Y. and Jun O. 2011. Retrieving Information about Real World Activities from the Web. IPSJ SIG technical reports 2011-MBL-59(2), 1-7, 2011-08-29
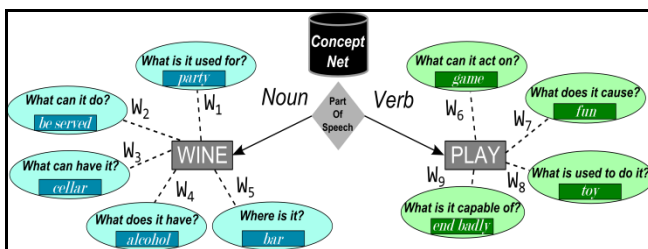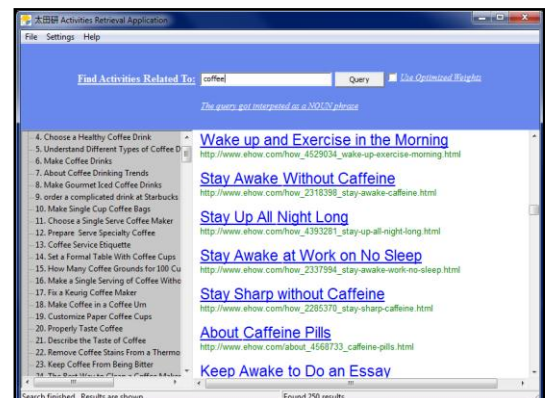
Fig. 1 ConceptNet and relation weights

Fig. 2 - The GUI of the system