

**Search Result Diversity Evaluation based on Intent Hierarchies**

**Abstract:**

Search result diversification aims at returning diversified document lists to cover different user intents of a query. Existing diversity measures assume that the intents of a query are disjoint, and do not consider their relationships. In this paper, we introduce intent hierarchies to model the relationships between intents, and present four weighing schemes. Based on intent hierarchies, we propose several hierarchical measures that take into account the relationships between intents. We demonstrate the feasibility of hierarchical measures by using a new test collection based on TREC Web Track 2009-2013 diversity test collections and by using NTCIR-11 IMine test collection. Our main experimental findings are: (1) Hierarchical measures are more discriminative and intuitive than existing measures. In terms of intuitiveness, it is preferable for hierarchical measures to use the whole intent hierarchies than to use only the leaf nodes; (2) The types of intent hierarchies used affect the discriminative power and intuitiveness of hierarchical measures. We suggest the best type of intent hierarchies to be used according to whether the nonuniform weights are available; (3) To measure the benefits of the diversification algorithms which use automatically mined hierarchical intents, it is important to use hierarchical measures instead of existing measures.

**Existing System:**

People tend to meet their daily information needs by issuing keywords into search engines. However, these keywords, i.e. queries, are often ambiguous or broad. The queries usually have several interpretations or aspects, also known

as subtopics or user intents. When users submit the same query to retrieval systems, they may want different information returned to fulfill their own information needs. This poses a challenge to search engines when the user intent cannot be known in advance.

**Proposed System:**

We propose four weighting schemes that are used to model the node weights in an intent hierarchy. We examine the impact of the different types of intent hierarchies, including whether the leaf nodes have the same depth and which weighting scheme is used, for hierarchical measures in terms of discriminative power and intuitiveness. In the experiments, we find the best type of intent hierarchies when nonuniform weights are available and when only uniform weights are known respectively.

We find that it is crucial for hierarchical diversification algorithms to be evaluated by hierarchical measures. The benefits in search result diversification by re-ranking

the results to cover the automatically generated hierarchical intents as much as possible may be invisible to existing measures that measure the diversity using intent lists. Hierarchical diversification algorithms show more gains when evaluated by hierarchical measures than existing measures.