Abstract

This paper proposes a query classification system for a one-click search system that uses feature vectors based on snippet similarity. The proposed system targets the NTCIR-10 1CLICK-2 query classification subtask and classifies queries in Japanese and English into eight predefined classes by using support vector machines (SVMs). In the NTCIR-9 and NTCIR-10 tasks, most participants used complex features or rules that depend strongly on language characteristics. The authors propose a new method that uses feature vectors created by using snippet similarities instead of the above mentioned features. In the proposed method, feature vectors have fewer dimensions, provide better generalization, lower language dependency, and reduced computer resources. This method achieved accuracies of 0.93 for a Japanese task and 0.91 for an English task.

References

- Alan H Cheetham and Joseph E Hazel. Binary (presence-absence) similarity
- Naoki Orii, Young-In Song, and Tetsuya Sakai. Microsoft research asia at the ntcir-9 1click task. Proceedings of NTCIR-9, 2011.

**Index Terms**

Computer Science  
Information Sciences

**Keywords**

Query Classification  
Dimension Reduction  
Intent  
Mobile