

Two human judges both majoring in Foreign Languages were asked to assess the results of question pattern extraction and give a label to each extracted question pattern. A pattern will be judged as “good” if it clearly expresses the answer preference of the question; otherwise, it is tagged as “bad.” The precision rate of extraction for these 200 questions is shown in Table 5. The second column indicates the precision rate when both of two judges agree that an extracted question pattern is “good.” In addition, the third column indicates the rate of those question patterns that are found to be “good” by either judge. The results imply that the proposed pattern extraction rules are general, since they are effective even for questions independent of the training and development data. Table 6 shows evaluation results for “two ‘good’ labels” of the first five questions. We summarize the reasons behind these bad patterns:

- Incorrect part-of-speech tagging and chunking
- Imperative questions such as “*Name the first private citizen to fly in space.*”
- Question patterns that are not specific enough

For instance, the system produces “*what name*” for “*What is the name of the chronic neurological autoimmune disease which ...?*”, while the judges suggested that “*what disease.*” Indeed, some of the patterns extracted can be modified to meet the goal of being more fine-grained and indicative of a preference to a specific type of proper nouns or terminology.

V. CONCLUSION

In this paper, we introduce a method for learning query transformations that improves the ability to retrieve passages with answers using the Web as corpus. The method involves question classification and query transformations using a learning-based approach. We also describe the experiment with over 3,000 questions indicates that satisfactory results were achieved. The experimental results show that the proposed method provides effective query expansion that potentially can lead to performance improvement for a question answering system.

References

- [1] Ittycheriah, A., Franz, M., Zhu, W.-J., and Rathaparkhi, A. 2000. IBM’s statistical question answering system. In Proceedings of the TREC-9 Question Answering Track, Gaithersburg, Maryland.
- [2] Hovy, E., Gerber, L., Hermjakob, U., Junk, M., and Lin, C.-Y. 2000. Question answering in WebClopedia. In Proceedings of the TREC-9 Question Answering Track, Gaithersburg, Maryland.
- [3] Moldovan D., Pasca M., Harabagiu S., & Surdeanu M. 2002. Performance Issues and error Analysis in an Open-Domain Question Answering System. In Proceedings of the 40th Annual Meeting of ACL, Philadelphia, Pennsylvania.
- [4] Hildebrandt, W., Katz, B., & Lin, J. 2004. Answering definition questions with multiple knowledge sources. In Proceedings of the 2004 Human Language Technology Conference and the North American Chapter of the Association for Computational.
- [5] Radev, D. R., Qi, H., Zheng, Z., Blair-Goldensohn, S., Fan, Z. Z. W., and Prager, J. M. 2001. Mining the web for answers to natural language questions. In Proceedings of the International Conference on Knowledge Management (CIKM-2001), Atlanta, Georgia.
- [6] Hermjakob, U., Echihiabi, A., and Marcu, D. 2002. Natural Language Based Reformulation Resource and Web Exploitation for Question Answering. In Proceeding of TREC-2002, Gaithersburg, Maryland.
- [7] Agichtein, E., Lawrence, S., and Gravano, L. Learning to find answers to questions on the Web. 2003. In ACM Transactions on Internet Technology (TOIT), 4(2):129-162. 8. Melamed, I. D. 1997. A Word-to-Word Model of Translational Equivalence. In Proceedings of the 35th Annual Meeting of ACL, Madrid, Spain.
- [8] Yi-Chia Wang, Jian-Cheng Wu, Tyne Liang, and Jason S. Chang. 2004. Using the Web as Corpus for Un-supervised Learning in Question Answering, Proceedings of Rocling 2004, Taiwan.
- [9] Google. (2008,July). *We knew the web was big* <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>.
- [10] Wikipedia.(2009,October). *Wikipedia*. <http://en.wikipedia.org/wiki/Wikipedia>;
- [11] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In *In 6th int'l semantic web conference, busan, korea* (pp. 11–15). Springer.