Are Web Search Queries Evolving into a Language of their own?

Rishiraj Saha Roy

Monojit Choudhury

Niloy Ganguly

IIT Kharagpur

Microsoft Research Lab India

IIT Kharagpur

ABSTRACT

It is believed that Web search queries are becoming more structurally complex over time. However, there has been no systematic study that quantifies such characteristics. In this work, we briefly discuss the hypothesis that queries are evolving into a unique linguistic system. We probe this hypothesis by examining the structure of Web queries from the first principles in the light of well-established methodologies from natural language understanding. We apply an unsupervised query segmentation algorithm which uses only query logs to identify the basic structural units of Web queries, and then we deduce two syntactic categories for query segments - content and intent. Next, for exploring corpus-level relationships with natural language text, we employ complex network modeling on query logs. Preliminary results of all these experiments show quantitative and qualitative proof that queries are not just some arbitrary form of text between random word sequences and natural language - they have distinct structural properties of their own. This research has potential impact on query understanding and language evolution research.