

Evaluating the coverage of entities in knowledge graphs behind general web search engines (Poster)

Yiming Zhao

Center for Studies of Information Resources, Wuhan University, Wuhan, China, 430072

E-mail: zhaoyiming@whu.edu.cn; Phone: +86-18942901531

Feicheng Ma

Center for Studies of Information Resources, Wuhan University, Wuhan, China, 430072

E-mail: fchma@whu.edu.cn; Phone: +86-13507119710

Xue Xia

School of Information Management, Wuhan University, Wuhan, China, 430072

E-mail: xiaxue_2014@foxmail.com; Phone: +86-18672779723

Web search engines, such as Google and Bing, are constantly employing results from knowledge organization and various visualization features to improve their search services. Knowledge graph, a large repository of structured knowledge represented by formal languages such as RDF (Resource Description Framework), is used to support entity search feature of Google and Bing (Demartini, 2016). When a user searches for an entity, such as a person, an organization, or a place in Google or Bing, it is likely that a knowledge card will be presented on the right side bar of the search engine result pages (SERPs). For example, when a user searches the entity *Benedict Cumberbatch* on Google, the knowledge card will show the basic structured information about this person, including his date of birth, height, spouse, parents, and his movies, etc.

The knowledge card, which is used to present the result of entity search, is generated from knowledge graphs. Therefore, the quality of knowledge graphs is essential to the performance of entity search. However, studies on the quality of knowledge graphs from the angle of entity coverage are scant in the literature.

This study aims to investigate the coverage of entities of knowledge graphs behind Google and Bing. The major research questions are: how is the coverage of entities in knowledge graphs behind Google and Bing in general and how is it in different domains according to the classification of Wikipedia?

Entities from Wikipedia were randomly selected to test the coverage of entities of knowledge graphs. According to the classification of Wikipedia, there are 12 major categories, including general reference, culture and the arts, geography and places, health and fitness, history and events, and so forth. We randomly selected 60 entities from each major category and 720 entities were collected in total.

Each entity was submitted to Google and Bing separately. If the search engine returns a knowledge card for an entity, it is regarded that this entity is covered by the knowledge graph of this search engine. The occurrences of knowledge cards in Google and Bing were counted and compared. A Chi-square test of independence was employed to test the difference of the coverage of entities between two search engines.

Google returned 259 knowledge cards for the 720 tested entities while Bing returned 385 cards. The result of the Chi-square test suggests that the coverage of entities in the knowledge graph is significantly different between Google and Bing (χ^2 (1, n= 1440) =45.667, $p=0.000<0.001$).

Then, the occurrence frequency of knowledge cards of Google and Bing was compared in 12 major categories under Wikipedia. A three-dimensional cross-tabs analysis was done to explore the relationships between categories of entity and the coverage rate of entities for Google and Bing. The number of returned entities in knowledge cards of Google and Bing is displayed in Figure 1.

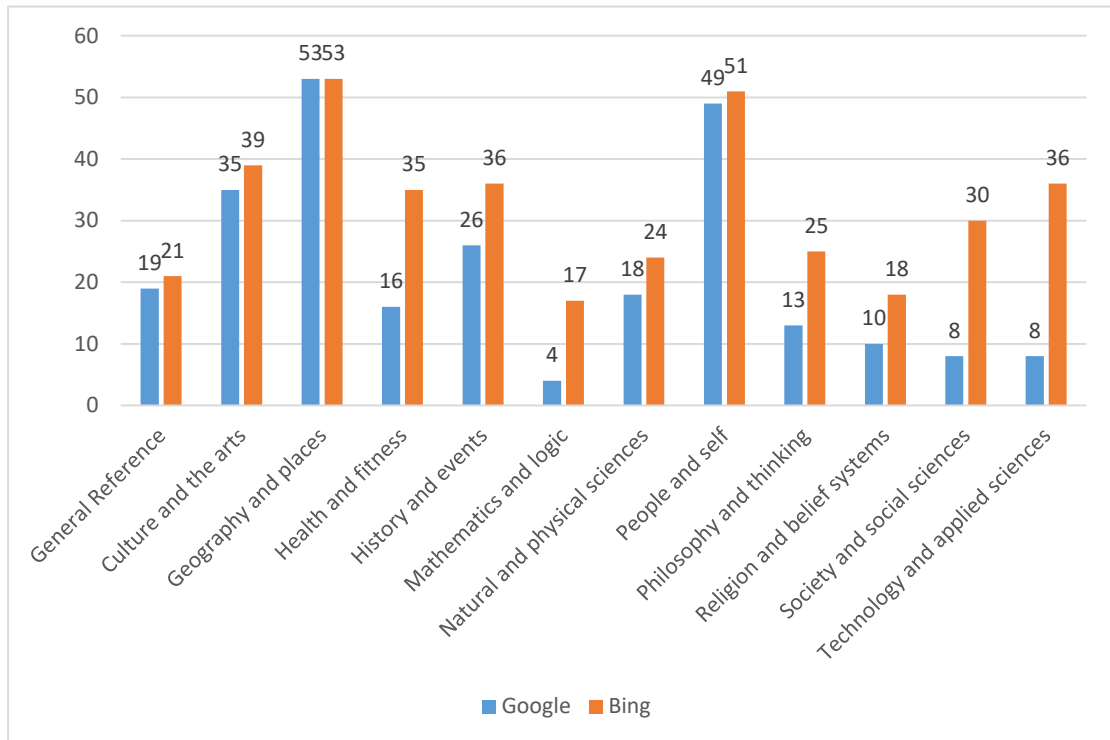


Figure 1. The number of returned entities in knowledge cards (the total number of testing entities in each category is 60)

This study has been completed and further discussion will be presented in the formal poster.

Reference

Gianluca Demartini. (2016). A Tutorial on Leveraging Knowledge Graphs for Web Search. In *Information Retrieval*. Pavel Braslavski et al. (Ed.): RuSSIR 2015, CCIS 573, pp.24-37.