07 Search suitable for various viewpoints

- "Pitarie": Picture book search with graph index based search -

Abstract

We propose a similarity search method for finding similar objects in a large-scale database. The search method is based on a graph index, where each vertex corresponds to a object and two vertices are connected by an edge when they satisfy a certain similarity condition. The graph index shows small-world behavior, that is, vertices can be reached from every other vertex by a small number of steps. Hence, searching the graph results in quick termination of the search process. Furthermore, since the graph index is constructed based on similarity between two objects, the search method is versatile and can be applied to wide variety of media such as text, images and audio. When applied to complex objects that are more than two media combined, such as picture books which consists of text and illustration, users can search from various viewpoints; users can find picture picture books that are not only similar in content but also similar in style of illustration.

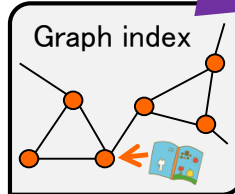## Picture book search system "Pitarie"

Search picture books from various viewpoints.

Case 1: Input a summary and search

Case 2: Search for books with similar contents

Case 3: Search for books with similar style of illustration (Example below)

Preparation of graph index in advance of search (Off line)

Picture books

Graph index

Graph index construction

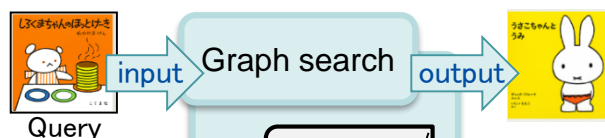Construct a graph by connecting similar objects, and use it as a index for search.

✔ "Small-world behavior" :
Any two objects are within a small number of edge hops.

✔ Since the graph construction is independent of media's characteristic, the proposed method is applicable to various media.

Fast search by utilizing graph index

Example You can find books with similar style of picture by various illustrators.

Query → input → Graph search → output → Similar book by a different illustrator.

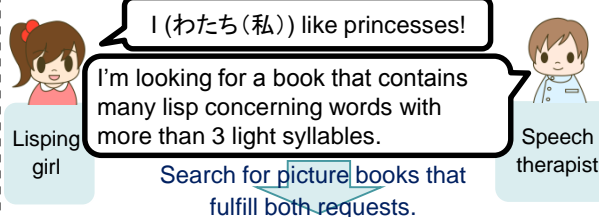Search for similar books by continuously traversing from a vertex (object) to its neighbors.

Cited picture books:
しろくまちゃんのほっとけーき、わかやまけん作、こぐま社、1972
うさこちゃんとうみ、ディックブルーナ作、福音館書店、1964
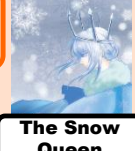(The cove-front-like illustration of "The snow queen" was in-house illustrated.)

Various applications

Example Search books suitable for articulation disorder training

I (わたち(私)) like princesses!

I'm looking for a book that contains many lisp concerning words with more than 3 light syllables.

Lisping girl

Speech therapist

Search for picture books that fulfill both requests.

Joint Research with a hospital.

The Snow Queen

phonetic histogram

Words in the book

References

[1] T. Hattori, T. Kobayashi, S. Fujita, Y. Okumura, K. Aoyama, "Pitarie: Picture Book Search with Interdisciplinary Approach," *NTT Technical Review*, vol.14, no.7, pp. 1-8, 2016.

[2] K. Aoyama, K. Saito, H. Sawada, N. Ueda, "Fast approximate similarity search based on degree-reduced neighborhood graphs," in Proc. *The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1055-1063, 2011.

Contact

Takashi Hattori Email: cs-liaison-ml at hco.ntt.co.jp

Learning and Intelligent Systems Research Group, Innovative Communication Laboratory

Innovative R&D by NTT
Open House 2019