Utilization of Large-Scale Academic Databases for Entity Identification

Integration of knowledge in academic/research community

"Information linkage" refers to a group of basic techniques for identifying information segments, or 'links', that refer to the same real-word entity or event. The extracted links can be utilized in various tasks such as information integration, structure analysis, and understandings of global structure of entities and events. We propose an overview of the issues and key techniques for information linkage and then, introduce our prototype linkage system currently under development.

Authority server for researcher information

- An authority server for researcher information based on researcher’s number in Kakenhi DB.
- Links the information with several information sources including a Web search engine and other databases.
- About 150,000 researchers’ information.
- Several technologies for extending functions of the server.

Linkage engine for bibliographic records

- Fast identification and reliable linkage based on word-based suffix array.

Example:

- Existing search engine
- Our linkage engine

The book itself cannot be identified.

- Specific record identified
- Identification based on free text phrase

Ranking document list that contains search terms.

- Link to the book itself!!

Adding information based on large-scale database.

Classifying records based on those connected entity.