Navigational Information as Emergent Intelligence of Spontaneously Structuring Web Space

Takashi Ishikawa

Department of Computer and Information Engineering, Nippon Institute of Technology, 4-1 Gakuendai, Miyashiro-town, 345-8501 Saitama, Japan
tisikawa@nit.ac.jp

Abstract. The paper describes emergent properties of spontaneously structuring web space such as collaborative editing encyclopedias and proposes a method to extract navigational information from the link structure of the web space. The navigational information consists of hyperlinks that are organized with approximate subsumption relations between sets of hyperlinks in the web pages. A case study for Wikipedia and Everything2 exemplifies the effectiveness of the proposed method that allows readers of free online encyclopedias to browse more efficiently.

1 Introduction

Web technologies have provided us a new medium of structuring knowledge in a flexible and efficient way. We call the medium hypertext, in which fragments of text are structured using hyperlinks that allow the reader to proceed to related fragments from the specified position of the text. A hyperlink is meta-information embedded in the text without affecting the contents. Whereas usual linear texts are structured with the contents itself. The author of a hypertext is able to restructure the text by changing the hyperlinks embedded in the text, but this flexibility cause the problem that the reader is hard to understand overall structure of the hypertext, that is so called glot in hyperspace phenomenon [10]. To overcome the problem authors of hypertexts provide secondary information such as summary, list and index, but they impede the efficiency of reading and disturb the thinking flow of the reader.

Collaborative authoring tools using web technology enable to describe many kinds of contents in the web pages as online encyclopedias [5]. A usual web page of the online encyclopedias is a list of entries or posts each of which...
is a hypertext including hyperlinks to connect to other web pages. The reader of an online encyclopedias selects an entry by the title or a hyperlink in other web page, reads the contents and proceeds to other entries in the same web page or other web pages by clicking an anchor text. The decision where to proceed is made by evaluating relevance of anchor texts to the interest of the reader in reading the entry. Therefore the reader has to read the whole contents of the entry to decide where to proceed. Contrastingly web pages in information sites such as Yahoo [23] have navigational information as category lists or local navigation menus so as to guide the readers in selecting anchor texts without reading the whole contents. The lack of navigational information in online encyclopedias is a deficiency of media to describe a lot of information or knowledge in the web pages.

The research hypothesis of the research is that there exists implicit navigational information as emergent intelligence of spontaneously structuring web space such as free online encyclopedias. In the hypothesis, navigational information means category lists or local navigation menus to guide the readers in selecting linked web pages. Looking navigational information is more efficient than reading whole contents in the web page, so it will become more efficient to use online encyclopedias as an information medium. Issues in verifying the hypothesis are that how the navigational information can be generated automatically from the contents and that how to add the navigational information to the web pages of online encyclopedias. An idea to approach the first issue is that the network structure of hyperlinks in web pages reflects the structure of information such as concept hierarchy so that the structural information must be extracted from the link structure [2, 3]. The second issue can be solved by implementing the method for extracting navigational information from link structure of online encyclopedias.

The aim of the research is to develop a computational method to extract navigational information from the link structure of online encyclopedias based on the research hypothesis. The navigational information means a subset of anchor texts in a web page to reduce items to be selected by the readers for the sake of search efficiency. The items in the subset are preferable being structured by the order of importance or relevance and selected so as not to reduce web pages that are reachable through the selected anchor texts. Therefore the problem to be solved is a kind of optimization problem to balance search efficiency and information amount.

2 Spontaneously Structuring Web Space

Collaborative free online encyclopedias such as Wikipedia [21] and Everything2 [6] are built with collaborative work in which the authors create and edit each own article with references to other articles. Articles in those encyclopedias are web pages that are created with web-based authoring system and references in the articles are anchor texts with hyperlink to other web