

Chapter 5

Acquiring the Gist of Social Network Service Threads via Comparison with Wikipedia

Akiyo Nadamoto

Konan University, Japan

Eiji Aramaki

The University of Tokyo, Japan

Takeshi Abekawa

National Institute Informatics, Japan

Yohei Murakami

National Institute of Information and Communications Technology, Japan

ABSTRACT

Internet-based social network services (SNSs) have grown increasingly popular and are producing a great amount of content. Multiple users freely post their comments in SNS threads, and extracting the gist of these comments can be difficult due to their complicated dialog. In this paper, the authors propose a system that explores this concept of the gist of an SNS thread by comparing it with Wikipedia. The granularity of information in an SNS thread differs from that in Wikipedia articles, which implies that the information in a thread may be related to different articles on Wikipedia. The authors extract target articles on Wikipedia based on its link graph. When an SNS thread is compared with Wikipedia, the focus is on the table of contents (TOC) of the relevant Wikipedia articles. The system uses a proposed coverage degree to compare the comments in a thread with the information in the TOC. If the coverage degree is higher, the Wikipedia paragraph becomes the gist of the thread.

INTRODUCTION

As Internet-based social network services (SNS's) have grown increasingly popular they are producing a great amount of content. An SNS consists of an Internet community, containing multiple

threads, with each thread containing comments posted by multiple users. It is difficult to obtain a gist of a thread because these comments are complicated. There are two types of SNS users. The first type becomes a member of the community and discusses the theme of the thread. We call this

DOI: 10.4018/978-1-4666-2026-1.ch005

user an “inside user.” The second type, an “outside user,” is a casual browser who simply views the SNS to acquire information from it.

Inside users can sometimes enter into heated discussion, which prompts them to concentrate on only one issue and lose track of the actual theme. When this happens, inside users may want to know how relevant their points are to the discussion. We therefore believe it would be beneficial to present these users with information to help them understand the gist of the discussion at a glance.

Outside users can theoretically obtain detailed information on the theme of a thread from the SNS, since community members who are inside users have good knowledge about the theme. Inside users, however, often do not explicitly provide basic information, because this is considered to be tacit knowledge, making it difficult for outside users to obtain. Moreover, a thread might contain many comments from outside users, making it difficult again to grasp the gist. It would be convenient for outside users to obtain the gist of a thread at a glance, while at the same time acquiring basic information about it.

In this paper, we propose a system that presents the gist of an SNS thread’s information by comparing the comments in the thread with Wikipedia article content. Wikipedia articles are posted by different users on the basis that a “neutral point of view is the fundamental principle of Wikipedia” (Wikipedia, 2011). We therefore consider a Wikipedia article on a given theme to be based on a general viewpoint. In the present study, we extract articles from Wikipedia and compare each comment in a thread in the community content with the smallest structure in the article’s table of contents (TOC). We consider the table of contents in the articles of Wikipedia to be the gist and the content of a paragraph on Wikipedia as basic information.

In this paper, we define a “target thread” as one from which the gist is to be extracted, and a “target article” as an article on Wikipedia that is to be compared with the target thread. The process

of extracting a gist and basic information is as follows (Figure 1 and Figure 2):

- Extract noisy comments from a target thread.
- Identify target articles on Wikipedia.
- Compare the target thread and target articles on the basis of the table of contents based on our coverage degree.
- Extract the gist of information about the target thread.

RELATED WORK

In the field of natural language processing (NLP), there is a great deal of research about content summarization. Mani (2001) wrote about automatic summarization. Most research summarizes the content by calculating the importance of the sentences and structure of the text. Tombros et al. (1998) describe the summarization of the results of information retrieval by using a query. Li et al. (2009) proposed a method for enhancing diversity, coverage, and balance in the summarization of content. Their viewpoints were similar to that of our approach. Some research has been done on creating summary sentences from Web pages (Sun et al., 2007; Wang et al., 2007; Delort et al., 2003) using NLP. It is however difficult to use the technique for extracting the gist of SNS threads because almost all comments are short and their conversation is confusing. It would be convenient if inside and outside users could extract this gist at a glance. This paper proposes a system that shows this gist of a thread’s information by comparing its comments with Wikipedia article content. To determine the differences between two Web pages, Lim and Ng (2001) proposed a semantic change-detection (SCD) algorithm for detecting semantic changes between two bodies of HTML data. Their approach was based on the transformation of HTML data from the two sources into trees and the removal of common edges between

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/chapter/acquiring-gist-social-network-service/71891?camid=4v1

This title is available in InfoSci-Books, InfoSci-Multimedia Technologies, Communications, Social Science, and Healthcare, Business Ethics, Data Analysis, and Decision Support, InfoSci-Media and Communication Science and Technology, InfoSci-Select. Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Enabling Cellular Device to Device Data Exchange on WISDOM 5G by Actuating Cooperative Communication Based on SMNAT

Rajarshi Sanyal and Ramjee Prasad (2014). *International Journal of Interdisciplinary Telecommunications and Networking* (pp. 37-59).

www.igi-global.com/article/enabling-cellular-device-to-device-data-exchange-on-wisdom-5g-by-actuating-cooperative-communication-based-on-smnat/124795?camid=4v1a

An Energy Aware Multi-Path Routing Algorithm for Mobile Ad Hoc Networks

S. Harous, M. Aldubai and Q. Nasir (2008). *International Journal of Business Data Communications and Networking* (pp. 58-75).

www.igi-global.com/article/energy-aware-multi-path-routing/1458?camid=4v1a

Wireless Communications: Is the Future Playing Out as Predicted?

William Webb (2011). *Interdisciplinary and Multidimensional Perspectives in Telecommunications and Networking: Emerging Findings* (pp. 9-18).

www.igi-global.com/chapter/wireless-communications-future-playing-out/52173?camid=4v1a

Application of Extrinsic Information Transfer Charts to Anticipate Turbo Code Behavior

Izabella Lokshina (2013). *Advancements and Innovations in Wireless Communications and Network Technologies* (pp. 97-104).

www.igi-global.com/chapter/application-extrinsic-information-transfer-charts/72419?camid=4v1a