AI & Soc
DOI 10.1007/s00146-013-0522-7

ORIGINAL ARTICLE

Tag line generating system using knowledge extracted from statistical analyses

Hiroaki Yamane · Masafumi Hagiwara

Received: 31 August 2012 / Accepted: 9 October 2013

Abstract This paper proposes a tag line generating system based on knowledge obtained by tag line analyses. Using a tag line corpus, two features of tag line—part-of-speech *N*-gram and word usage—are obtained as knowledge. Based on these features, the tag line generating system generates tag line candidates. More specifically, for a given theme, a keyword is entered into the proposed system. For each keyword, the system selects related terms unique to the theme. And enormous number of candidates is generated by using a large-scale *N*-gram corpus. The proposed system subsequently selects candidates based on the degree of similarity, grammatical structures, and mutual information volume. The highest scored candidates are selected as final output. The performance of the proposed system has been evaluated by subjects in terms of sentence quality, category appropriateness, keyword appropriateness, and overall quality. The experimental results show that the proposed system is able to generate suitable tag lines for a given theme and keywords, indicating a potential online tag line generator in the future.

Keywords Tag line · Statistical analysis · Sentence generation

H. Yamane (✉) · M. Hagiwara
Department of Information and Computer Science, Keio University, Yokohama 223-8522, Japan
e-mail: yamane@soft.ics.keio.ac.jp

M. Hagiwara
e-mail: hagiwara@soft.ics.keio.ac.jp

1 Introduction

Most human knowledge and communication are expressed linguistically. Written languages are especially influential. For example, "tag lines" play an important role in the field of advertising and must be capable of appealing to people in a short amount of time. Tag lines attract people by expressing the uniqueness and merits of an object briefly and effectively with short sentences.

In the field of advertising, there are a number of studies that analyze catchphrases, slogans, and tag lines (Kohli et al. 2007; Kitamura et al. 1981). There are also advertising slogan generators (http://www.sloganizer.net; THE-PCMAN-WEBSITE) in which the keywords are in a fixed line. As for sentence generation, Banko et al. (2000) proposed a system that produces headlines based on statistical translation.

In Japan, various tag line related systems have already been proposed. Matsudaira and Hagiwara (2004, 2005) introduced tag line making support systems using genetic programming. Nishihara et al. (2007) proposed algorithms to extract attractive research titles for workshops. They applied the pointwise mutual information (PMI) algorithm for evaluating understandability and amusingness of the titles. Nakano and Onisawa (2007) proposed a tag line-generating system using blogs. These systems, however, do not consider the uniqueness of tag lines from a statistical point of view. Therefore, it is necessary to investigate a number of tag lines in order to establish uniqueness.

In this paper, we utilize a tag line encyclopedia book (Kuno 2008) to obtain a multitude of examples of tag lines and apply their features to generate sophisticated tag lines. This book lists more than 6,400 tag lines and covers twelve fields such as food and livingware. Our proposed system uses Japanese Google *N*-gram Corpus (Kudo and Kazawa

Published online: 11 December 2013