ELSEVIER

Expert Systems with Applications 26 (2004) 575–582

Expert Systems with Applications

www.elsevier.com/locate/eswa

Business information extraction from semi-structured webpages

Nahk Hyun Sung[a,*], Yong Sik Chang[b]

[a]*Department of Management Information Systems, Yong-In University, 470 Samga-dong, Yongin, Kyungki 449-714, South Korea*
[b]*Department of e-Business, Hanshin University, 411 Yangsan-dong, Osan, Kyungki 447-791, South Korea*

Received 9 October 2003; accepted 1 December 2003

Abstract

To protect online consumers, as OECD Guidelines recommend, Internet shopping malls should provide information about their business on their webpages. In Korea, The Consumer Protection Law in Electronic Commerce, forced Internet shopping malls to provide their business information, so that consumers could easily identify them. Since most Korean Internet shopping malls provide consumers with business information in a semi-structured format on their homepages, a software agent can easily identify them.

To investigate automatically the provision of the business information with the Internet shopping malls, this article proposes the methods of gathering URLs of Internet shopping malls, of monitoring alterations of webpages, and of extracting business information. Business information extraction in our research is based on synonyms and indicator words of the attributes. We used inductive learning to raise the efficiency of information extraction. With experiments, we showed the potentialities of our agent system. The average extraction accuracy of our agent system was 89.3%.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Electronic commerce; Internet shopping mall; Business information; Information extraction; Agent

1. Introduction

The factors that affect public confidence in Internet shopping malls are reputations of shopping malls, clearness of business information, protection policies for consumers' privacy information, and security policies for payment, etc. Among these factors, clearness of business information is a basic factor that can lead to confidence in shopping malls in the electronic commerce environment.

OECD announced the Guidelines for Consumer Protection in Electronic Commerce in 1999 (OECD, 1999). The OECD Guidelines and the Guidelines of Membership Nations, created shortly thereafter, specify that Internet shopping malls should provide at least a basic minimum of business information on their webpages, including the name of the business, the name of the representative, geographical address, telephone number, fax number, e-mail address, and business license number.

As examples, Fig. 1 depicts two homepages including business information: BEST BUY Co., INC. (www.bestbuy.com) in the US and LGeshop (www.lgeshop.com) in Korea.

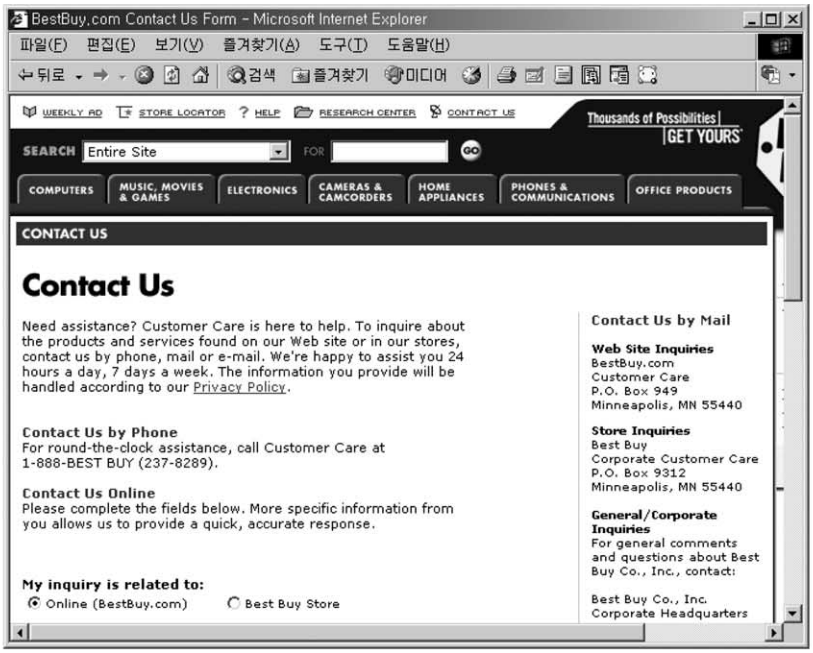
As we can see in the examples of BEST BUY Co., Inc. and LGeshop, while most Internet shopping malls in the US provide their business information, scattered in several pages, in an unstructured format, most Korean Internet shopping malls provide their business information on the bottom of their homepages in a semi-structured format. In Korea, The Consumer Protection Law in Electronic Commerce, which came into effect in July 2002, forced Internet shopping malls to provide a minimum of seven forms of business information, including the name of the business, the name of the representative, geographical address, telephone number, fax number, e-mail address, and business license number, so that consumers could easily identify them.

Therefore, in Korea, Internet shopping malls should provide their business information on their webpages. Since most Korean Internet shopping malls provide their business information in a semi-structured format, an agent can easily identify them when compared with other countries such as the US.

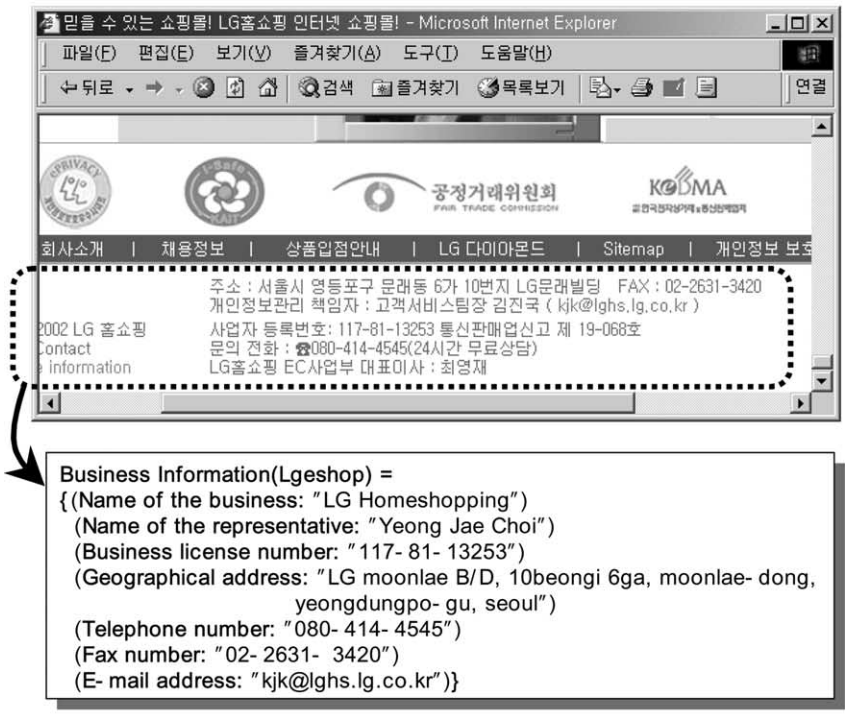
If any shopping mall intentionally omits all or a part of the required business information, they can be regarded

* Corresponding author. Tel.: +82-1620-71809; fax: +82-3133-02885.
E-mail addresses: nhsung@yongin.ac.kr (N.H. Sung); yschang@hanshin.ac.kr (Y.S. Chang).

0957-4174/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2003.12.008



(a) BEST BUY Co., INC.(www.bestbuy.com)



(b) LGeshop(www.lgeshop.co.kr)

Fig. 1. Examples of homepages providing with business information. BEST BUY Co., Inc. in the US provides its business information in an unstructured format and LGeshop in Korea provides its business information, given inside the rounded rectangle made of dotted lines, in a semi-structured format. It was translated into English for readability.

as a suspect of online fraud. If an organization would detect Internet shopping malls which lack business information and admonish them for not providing business information, it would enhance public confidence in electronic commerce; however, it is difficult for a person to visit a large number of Internet shopping malls' homepages to investigate whether or not they provide business information.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات