

루씬 기반의 최저가 상품 검색 시스템 설계

김아용 · 정대진 · 계민석 · 김창수 · 정희경

배재대학교 컴퓨터공학과

Design of Search System Based on Lucene for Minimum Price Products

A-Yong Kim · Dae-Jin Jeong · Min-Suk Gye · Chang-Su Kim · Hoe-kyung Jung

Department of Computer Engineering, PaiChai University

E-mail : janlssary@naver.com, jdj@realtimetech.co.kr, iam7ms@gmail.com,

{MIE-ddoja,hkjung}@pcu.ac.kr

요 약

인터넷 이용률 증가와 스마트 기기의 대중화로 인해 소비자는 매장에서 구매하던 비용을 온라인 쇼핑 시장으로 전환하고 있다. 이로 인해 사용자의 소비패턴과 소비문화도 변화하고 있다. 오픈 마켓은 웹과 모바일을 통해 유통 채널을 확장하고 소비자의 유치를 위해 다양한 이벤트와 최저가 정책, 안전 거래 등을 제공한다.

본 논문에서는 오픈 마켓에서 판매하는 상품의 정보를 수집하고 분석하여 사용자에게 최저가 상품 정보를 제공하는 검색 시스템을 설계한다.

ABSTRACT

Has been switched to the online shopping market in stores of the consumer is from increased utilization and smart devices, the internet popularization. That is why has been converting the user's consumption patterns and consumer culture. Open markets is provides of making a wide variety of events and lowest price policies, safe transactions etc, for attract the consumers of expand distribution channels of the web and via mobile.

In this paper, a designs of provides a search system for minimum price product information to the user of Information collect and analyze on sale from open market.

키워드

Hadoop, Lucene, Nutch, Solr, Ubuntu

1. 서 론

인터넷 기술 발전과 스마트의 대중화는 매장에서 직접 구매하던 소비자들이 온라인 쇼핑 시장으로 전환되고 있다[1]. 그리고 오픈 마켓은 기존의 저가 전략에서 차별화 전략으로 전환하여 소비자를 유치하고 있다. 또한, 모바일 앱으로 거래 시 새로운 혜택을 부여하여 소비자의 모바일 사용을 장려하고 있다[2]. 하지만 다양한 오픈 마켓과 많은 종류의 상품이 존재하기 때문에 소비자는 모든 사이트의 상품 가격을 비교하기는 힘든 상황이다[3].

본 논문에서는 루씬과 KMP(Knuth Morris Partt) 알고리즘을 사용하여 오픈 마켓에 등록되어 있는 상품 가격과 각각의 사이트마다 차별화된 혜택들을 수집하고 비교하는 최저가 상품 검색 시스템을 설계한다.

II. 관련 연구

본 장에서는 설계하는 시스템에서 사용하는 루씬과 KMP 알고리즘에 대해 기술한다.

2.1 루씬

루씬은 자바로 개발된 강력한 텍스트 검색 엔진 라이브러리이며 크로스 플랫폼을 필요로 하는 거의 모든 응용 프로그램에 적합한 기술이다[4]. 그리고 루씬은 웹 크롤링과 문서 필터링, 실시간 서버, 사용자 인터페이스 등 검색 엔진 구성도 쉽게 구현할 수 있다. 또한, 다양한 언어(C/C++, C#, Ruby, Perl, Python, PHP 등)를 지원하여 많은 웹 사이트에서 사용하고 있다.

2.2 KMP

KMP 알고리즘은 Knuth와 Morris, Pratt에 의해 발견되었다[5]. KMP 알고리즘은 기존의 문자열 매칭 알고리즘이 문자열을 읽는 과정에서 발생하는 정보의 낭비를 방지함으로써 실행시간을 단축시켰다. 그리고 KMP 알고리즘은 기존 문자열 매칭 알고리즘에 비해 단순하다. 또한, KMP 알고리즘은 입력 문자열에 대해 패턴의 비교 수를 최소화하여 효율적인 알고리즘을 구현할 수 있다.

III. 시스템 설계

제안하는 최저가 상품 검색 시스템의 구조는 그림 1과 같이 세 부분으로 구성되어 있다. 첫 번째의 Crawler는 웹에 존재하는 각각의 오픈마켓의 상품명과 상품 가격, 혜택들을 수집한다. 두 번째로는 수집된 Crawldb와 LinkDB를 색인하는 Indexer가 있다. 마지막으로 사용자의 키워드를 처리하는 Searcher로 분류된다.

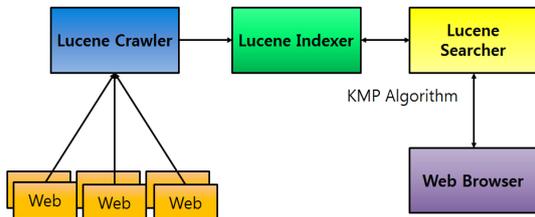


그림 1. 최저가 상품 검색 시스템 구조도

Searcher는 KMP 알고리즘을 사용하여 사용자가 입력한 키워드를 분석하여 그림 2와 같이 웹 브라우저를 통해 사용자에게 검색 결과를 출력한다.

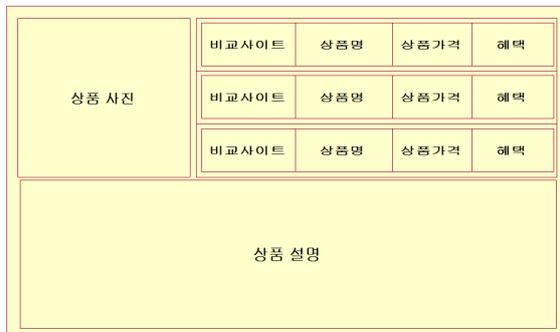


그림 2. 최저가 상품 검색 시스템의 검색 화면

상품 가격 부분에서는 상품 가격을 오름차순으로 정렬하여 출력하며 상품이 등록되어 있는 사이트명과 상품명, 상품가격 그리고 혜택 순으로 출력한다.

혜택은 오픈마켓마다 차별화를 두어 각각 다르다. 그로인해 각각의 오픈마켓에서 제공하는 혜택을 미리 정의해야 한다. 또한, 상품마다 혜택을 다르게 적용하기 때문에 그림 3과 같이 사용자가 검색한 상품의 혜택 유무를 분석하는 과정이 필요하다.

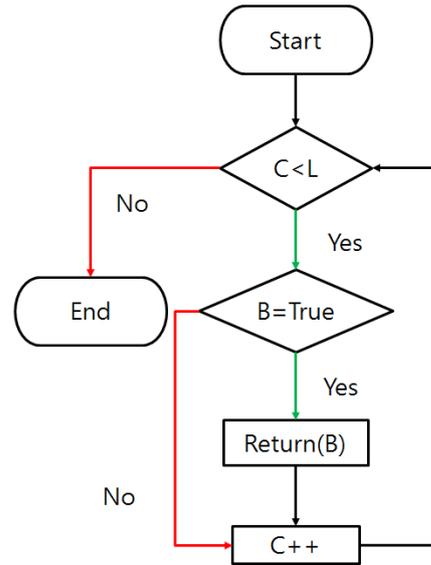


그림 3. 상품 혜택을 분석하는 순서도

IV. 결론

장기 불황으로 인해 소비자의 소비문화가 오프라인 시장에서 온라인 시장으로 이동하고 있다. 또한, IT 기술 발전과 스마트 기기의 사용 증가로 인해 온라인 시장의 성장을 가속화하고 있다.

오픈마켓들은 소비자 유치를 위해 모바일 앱과 상품 가격, 혜택들을 차별화하고 있다. 하지만 많은 오픈마켓들이 등장하여 소비자들은 오픈마켓에서 제공하는 정보들을 인식하기에는 어려움을 겪고 있다.

본 논문에서는 오픈마켓에서 판매하는 상품명과 상품 가격, 혜택들을 수집하여 비교하는 최저가 상품 검색 시스템을 설계했다. 설계한 시스템은 최저가의 상품을 기준으로 정렬하여 출력하며 해당 사이트의 혜택 등을 출력하여 사용자들이 유용하게 사용할 것이라고 예상된다.

참고 문헌

[1] 정수연, 박철, “e-비즈니스 일반: 온라인쇼핑 가치가 사이트속성중요도와 쇼핑의도에 미치는 영향에 관한 연구-한국과 영국의 차이를

- 중심으로, “ e-비즈니스연구, pp.69-94, 2008
- [2] 김창수, 주희엽, “모바일 쇼핑의 고객만족에 영향을 미치는 가치인식과 긍정적 사용의도에 관한 연구,“ 한국정보기술학회논문지, pp.141-148, 2014
 - [3] 이상철, 김상욱, 박선주, “가격 비교 서비스에서 상품 그룹 분석을 통한 추천 기법,“ 정보과학회논문지, pp.541-546, 2013
 - [4] McCandless, Michael, Erik Hatcher, and Otis Gospodnetic, “Lucene in Action: Covers Apache Lucene 3.0,“ Manning Publications, 2010
 - [5] Justin Kourie, Bruce Watson, and Loek Cleophas, “On Compile Time Knuth-Morris-Pratt Precomputation,“ Stringology, 2011