TELKOMNIKA, Vol.16, No.1, February 2018, pp. 455~462
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v16i1.5943

Designing an Agent for Information Extraction from Persian E-shops

Nasrin Rasouli, Leila Abedi*, Sara Ghaei
Department of Computer, Flavarjan branch, Islamic Azad University, Isfahan, Iran
Teacher Research Center, Computer Group, Isfahan, Iran
*Corresponding author, e-mail: abedi.leila@yahoo.com

Abstract

E-shops are among the most conventional applications of Electronic Commerce. In these shops, the buyers search for their goods through key words or classifications and read the product description provided by the sellers. Though, when the number of items is high, this gets to be difficult for the users. On the one hand, there are too many e-shops, and browsing in these shops to find the best and most appropriate goods is a difficult and time-consuming process. On the other hand, product descriptions are not the same in different websites, and there are different product forms. This study investigates about products and sellers in various websites based on the conditions and user requirements through software agents which present the extracted information in the form of a table to the users which enables them to compare prices and each seller's conditions without spending too much time for browsing. Using this method increases precision and recall indices comparing to a conventional user browsing

Keywords*: e-shops, intelligent agents, information extraction, persian language, ontology*

1. Introduction

Today, with the advent of the internet, e-commerce has become and extensive procedure in the era of information in which three dimensions of commerce i.e. the exchanged product/service, the sales procedure, delivery and customer services, have undergone some changes from their tangible physical mode to an electronic one. The variety in the degree of combining physical and electronic dimensions of commerce determines the level of electronic commerce. In case all these three dimensions are offered electronically, the highest level of e-commerce is offered. While in traditional commerce, all three levels are physical and tangible, e-commerce can be accordingly applied in all or some levels in the cycle of commerce which comprises of: searching for goods or services in accordance with needs and requirements; searching for suppliers and doing the negotiations, placing the order, delivery and making the payment, providing post-sales activities and customer services like guarantee [1].

In traditional transactions, the buyer must put a lot of time and effort to interpret the acquired data about goods and services, make the most optimized purchase decision, and finally pursue the negotiations, transaction and payment procedures. E-commerce seeks to minimize the buyer/seller's physical presence and activity in all purchase and sale stages and optimize this procedure [2]. A function of intelligent agents is making commercial procedures intelligent. After getting the initial data from the user, these software agents perceive the activity type and proceed the commercial procedure independently. Intelligent software agents can be utilized in a wide range of applications such as: e-mail, auctions, e-commerce control and supervision, and facilitating the procedures on the part of the buyer [3]-[4]. An application of software agents in e-commerce is investigating the buyers' shopping behavior model and then offering some suggestions such as presenting exclusive advertisements, customer's favorite products and directions to select the appropriate supplier, etc. Paying attention to each buyer's shopping procedure and saving his/her activities such as clicks, searches, reading catalogs, etc. provides the seller with valuable data about the buyer's interests and priorities, and provides the opportunity to investigate and analyze these issues for the next time that the buyer refers to the e-shop so that to specify and personalize the shop's atmosphere in accordance with his/her needs and priorities not only to create a particular interest and relation with the customer, but also to optimize the shopping procedure for him/her in a short time [5]. Intelligent software

Received March 3, 2017; Revised January 18, 2018; Accepted January 26, 2018

agents are beneficial in all stages from the most primitive parts of the procedure, i.e. saving personal information about each individual to investigate and analyze and finally offer a personalized environment.

E-shops are among the most conventional functions of e-commerce. In these shops, users search for goods using keywords or classifications and read the item description provided by sellers. Though when the number of goods is high, it gets difficult for the buyer. On the other hand, the number of e-shops is so high. Browsing in these shops to find the best and most appropriate product is so difficult and time-consuming. On the other hand, item descriptions are not the same in different websites, and there are different templates for the products; for instance, item description in Digikala and Rayan Saba websites is as in Figure 1 and Figure 2.

Figure 1. Item Description in Rayan Saba

Figure 2. Item Description in Digikala

Item description are presented in various templates in different websites. This study seeks to design an agent to extract information about the required goods from different websites

and offer product shopping suggestions based on the user's interest and specifications. Therefore, without searching in various e-shops, the user can easily buy the item that matches his/her specifications. In the following, Research Background is presented in section 2, Information Extraction in section 3 and Model Architecture in section 4 and in the end the Research Conclusion.

2. Background

Many studies have focused on the application of intelligent agents in electronic sales websites. Far, et al presented a method to prevent agents' random behavior. Working with multiple agents in an online sales system requires many scenarios which change into the finite state machine (FSM) to make controlling agents' behavior possible and prevent their random behavior [6]. Mezei et al presented a protocol to allocate tasks to sales agents in wireless sensors. In most current methods of task allocation, relational expenses are rarely taken into account. This article presents a protocol to decrease the number of the massages sent for task allocation to save energy in the network [7]. Reverse auction of many features are used to centralize large organizations. Dingwei et al applied feature-based grouping to evaluate sales, which is necessary for assessing fairness and resource allocation method [8]. Wang et al investigated the possibility to use mobile agents in mobile devices. Moreover, a sales agent architecture named J-phone is implemented for Mobile commerce which helps the users in supervising sales conditions and making decisions in multiple sales websites [9]. Sandu et al presented server architecture based on the distributed agent of English action. Auction is managed by a hierarchical structure of agents distributed on the network. The results of this method show considerable improvement in server efficiency comparing to when one agent is used to manage the sale [10]. In a multi agent autonomous and intelligent system, the agents need to work together and focus on their conventional and personal purposes in an environment with limited or very few resources. Figuring out how these resources are efficiently allocated is crucial. Li and Ma presented a mathematical perspective to solve this problem [11]. In electronic commerce, agent-based autonomous online marketplace is a developing research study. Most studies are focused on sales mechanisms and strategies while research on other key subjects in constructing autonomous online marketplace system like sales positioning and collaboration among commercial partners are neglected. Huang and Youliu proposed a solution to this problem based on the mobile agent technology and game theory [12]. As a result of extensive use of the internet, electronic marketplace has become so conventional. The internet provides complete information about the market and an infrastructure to run marketplaces with lower executive expenses. Descending bid and the second highest bid are the most conventional forms of electronic auctions. Akkaya and Badur presented a dynamic model of electronic marketplace to investigate how customer satisfaction is influenced by various types of descending bids. This subject is theoretically investigated in economics about various methods of static sales. To overcome the limitations of this perspective, a new agent-based model is presented in which the researchers have utilized a simulator to investigate the behavior and interaction of autonomous agents in socio-economical environments [13]. In group shopping, there are two different roles: provider and sales agent. The major relation of the agents is created among providers and sales agent. To analyze this relationship, Qian initially explained sales procedure and investigated the conflict in objectives and asymmetry in information in this agent-oriented relationship and then presented a solution to this problem [14]. Recent studies indicate that web-based educational environment and active learning can improve learning efficiency. Cheung developed a system in which effective learning is created by a web-based system, and learners' active involvement is facilitated through game competition for electronic commerce subjects and sales agent programming. Game competition is about a sale among mobile agents to have auction about their customers' resources. Mobile agents are in fact software agents which are programmed by learners [15]. Sales mechanism is extensively utilized in web-based sites; though they might prove less efficient in the future, and there might be a need to revolutionize the sales agents that are compatible with dynamic sales environment. Cheung has utilized genetic algorithm in sales agents. Cheung's proposed model helps the agents in strategy development through buying more goods with less price. Moreover, genetic algorithm programming leads to perceiving proper strategy in the current situation [16]. A vital ability of intelligent agents is making rational, accurate and quick decisions in dynamic

environment in a rational period of time. Mesbah and Taghiyar introduced a new classification method based on positive and negative patterns. For this classification, data log history of TAC/AD sales is used. Agents that are equipped with classifiers can gain higher profit comparing to those that are not so [17]. Chen has designed and implemented a multi user and multi access marketplace in which the users can access the system through the web, devices equipped with wireless applicable protocol (WAP) and agents. Electronic marketplace supports a variety of auctions including English auction, Dutch auction, American auction, hidden price auction and mutual auction [18]. Distinctive pricing rule or pay as bid to replace with uniform pricing rules is offered in electronic markets. It is expected that this model lowers the prices in the market and decreases price instability. Using multi-agent perspective, Xiong and Okuma compared pay as bid and uniform pricing where each compatible agent presents bid prices based on Q-learning algorithm. The experimental results indicated that pay as bid leads to decrease in market prices and price instability [19].

3. Information Extraction

Information Extraction is a sort of information retrieval with the purpose of extracting the data which has template out of semi-structured on unstructured documents. Information Extraction is considered a subdivision of Natural Language Processing (NLP). By web information extraction, we mean recognizing and extracting the required items from web pages, which also provides the opportunity to collect data and information from some sources (websites and web pages) to create added value services such as collecting web data in accordance with customer needs to compare products while shopping, advanced searches and etc.

As to the rise in the amount of data in unstructured web pages, web information extraction is highly important. Software agents need to extract data to be able to process unstructured data based on the extracted information. Information Extraction Systems used to previously apply Natural Language Processing techniques such as grammar and lexicon while web information extraction systems apply machine learning and pattern mining methods in the template of webpages. Linguistic analysis implemented for unstructured texts cannot extract the existing HTML or XML tags in an online text. Hence linguistic analyses are not commonly utilized. To extract online data, other methods are applied [20-24].

Information extraction methods fall into two categories: wrapper-based information extraction and Information extraction based on conceptual model.

3.1. Information Extraction Based on Wrapper

Wrapper is a set of high precision rules extracting the content of a specific page. This program extracts the data related to each required item from webpages and puts them in a database. There has been various activities in the scope of creating wrappers, which can generally be divided into five categorized: methods based on inquiry language, methods based on Natural Language Processing, HTML structure-aware methods, deduction-based methods, and ontology based methods.

a. Methods Based on Inquiry Language: one of the initial steps to create wrappers is creating a specific language to retrieve data items from webpages. These languages are much simpler than general-purpose languages and help producers in creating wrapper as an instrument. In fact, it can be said that this method is totally done manually. The user must be aware of the tree structure of HTML tags and place of data items. A disadvantage of this method is the need for much human intervention in finding the existing data items (for each webpage). Another demerit is the fact the wrappers are so dependent on the structure of the pages. In case there are any structural or conceptual changes in webpages, executing these wrappers will accompany some errors.
b. Methods Based on Natural Language Processing: this method is conventionally applied in unstructured texts (free texts written in natural language)
c. HTML Structure-Aware Methods: these methods utilize the innate and structural specifications and properties of HTML documentations. In this method, before information extraction is carried out, a decomposition tree is created for an HTML file, and its labels are created and shown in the memory in a hierarchical way (in the shape of a tree). Then the extraction rules which are based on the displayed properties of the data are executed either automatically or semi-automatically.

d. Wrapper Deduction-Based Methods: this category of methods receives a set of educational samples and creates the extraction rules based on them. The major difference of this category of methods with methods of natural language processing is that, here, there is no dependency to the language- related conditions of the data. But they depend on those template specifications that implicitly present the structural properties of some parts of the received data.
e. Ontology-Based Methods: most previous methods relied on the structure of the items and how data items are displayed, based on which the extracted patterns and rules were specified. Though in addition to the display structure of the data, the data items can also be extracted relying on the data itself. To do so, an ontology is created for a specific domain by which the existing constants in a webpage are recognized, and based on them, the existing objects are extracted.

3.2. Information Extraction Based on Conceptual Model

Information extraction methods which are based on the conceptual model are applied for free texts. These types of data are extracted based on their grammatical aspects [25].

4. The Architecture of the Proposed Model

The model of customer's purchase model includes six stages of purchase procedure elaborated in the following. It is also attempted to investigate the function of intelligent software agents [26]-[27].

a. Need Recognition: this stage clarifies that the buyer has become aware of some of his/her hidden needs by receiving information and being exposed to advertisements, and is tempted to do the shopping [26].
b. Brokerage:
    a) Product Brokerage: when the need to buy is created in the buyer, based on the assessment of the data acquired about the product, the buyer must determine what she/he wants to buy. At the moment a large number of intelligent software agents are active in this stage in the internet, which help the customer through browsing and presenting a variety of a given product, brands, prices and other existing specifications in a shop or various shops to select the product. The outcome of this stage is achieving a set of products [26].
    b) Broker: this stage combines the set acquired from previous stages with each seller's facilities to help making decisions about where to buy. The fault in the previous stage i.e. merely concentrating on product brokerage is that product properties are not the only important things for the customer. Other criteria such as post sales customer services like guarantee, easy access to the product, delivery time and its expenses, raises and discounts and etc. are also influential in selection procedure. Sufficing to product specifications does not guarantee customer's favorable conditions about delivery, guarantee and etc. [3].
c. Negotiation: in this stage, price and other items of the contract are defined. Traditional commercial negotiations impose huge costs to both customer and seller parties. They also accompany other barriers like time limitation, possibility of leading to no result, physical presence and etc. In digital world, none of these are encountered [3], [28], [26].
d. Payment and Delivery: this stage can occur right after finishing the stage of negotiation or a while later. As mentioned earlier, in some cases, easy payment or appropriate delivery conditions can influence the stage of product brokerage and broker.
e. Product Services and Evaluation: this post purchase stage includes offering post-sales customer services and the investigation and evaluation of overall satisfaction about the shopping and decision making experience. After evaluating satisfaction, the most important role of using agents in the stage of evaluation is supporting, maintaining and then improving customer satisfaction. Customer satisfaction management through raise and encouraging services to increase customer loyalty to the shop can be very influential in the profitability of electronic commerce [26].

The agent used in this model is a product and seller/selling brokerage type which help the user in finding the product and the intended seller based on the required property or specification. The framework of the proposed model is presented in Figure 3.

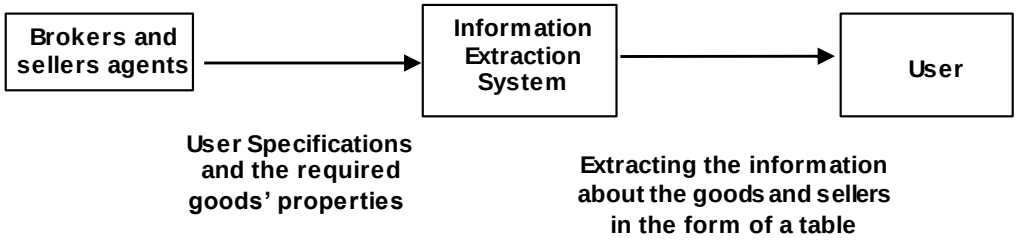

Figure 3. The Framework of the Proposed Model

Information extracting systems are also comprised of various components such as Document Retriever Medium which undertakes the task of collecting the contents of webpages. This component receives the address of a webpage as the input and then reads its content from the internet and returns it as output. Most webpages include meaningful data with formatting information. A webpage includes so many tags specifying how a webpage must be shown in a browser. Though these tags don't offer any further information about the concept of what is being shown. Therefore, Content Filter Module attempts to omit HTML tags. Content Filter Module's output which is a flat text is sent to Value Recognizer Module. This module plays a key role in extracting information and its major duty is to exert the rules of recognition that are related to the ontology of extraction and finding a set of candidates for extraction. As for each part of the webpage, there is a possibility to find some extracting ontology candidates, in case two or more candidates are found, it must be clarified that to which part of the ontology that part is related. One of the most important components of this system is Value Mapper Module. The main duty of this module is mapping candidate values to ontology components. The duty of this module is exchanging literal values to objects (a sample of the existing concepts in ontology). The output of this module, a data sample, would include a set of objects and the relation among them. Finally, Ontology Writer Module implements the data extracted outputs in the form of sample ontology.

For this system to work accurately, the initial ontology must be created accurately with high precision. We used Protégé software to create the initial ontology. This ontology is about computer and its accessories. Using this initial ontology, the users can search for a variety of computers and their accessories in different shops. In this model, if the initial ontology is adequately expressive and explanatory, the extracting procedure can be done completely automatically. In addition to the fact that the wrappers created in this way are resistant to the structural changes in the documentations. In fact these advantages are the result of using ontologies in recognizing data items shown in Figure 4.

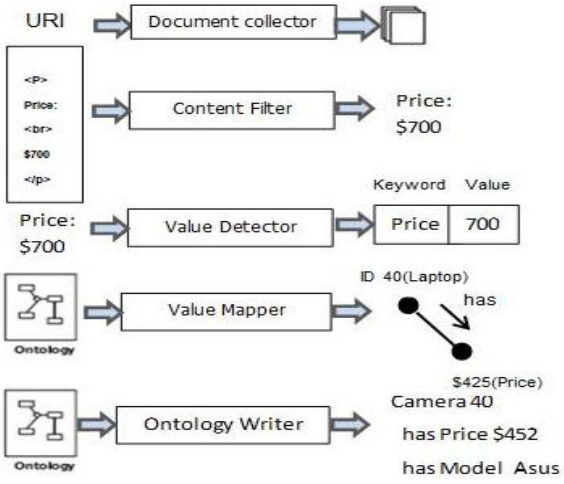

Figure 4. Information Extraction System

4.1. Model Evaluation

In data marketing systems Precision and Recall indices are utilized for evaluation; they are accordingly defined as follows:

$$\text{Precision}=\frac{tp}{tp+fp}$$

$$\text{Recall}=\frac{tp}{tp+fn}$$

Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. To evaluate the proposed model, ten users commonly searched for their required goods on electronic shopping websites. The requests of these ten users were also done simultaneously by software agents. The results obtained by software agents were closer to users' needs on the one hand, and were obtained in a much shorter time, on the other hand. The precision in users' browsing was previously 70% which increased to 90% by using software agents. Moreover, the index of recall in users' browsing was previously 50% which increased to 70% by using software agents.

5. Conclusion

Developing computer systems and extending the usage of information technology in today's life has led to the extreme supremacy of information such that this era is called information era. The amount of data and the degree of data usage are two fundamental indices to develop countries, the more the information volume increases, the harder its control and management would be. Hence merely data production and existence is not enough. Some instruments must be provided to use this volume of data. In fact the users must know how to respond to their need for information. As a result, information extraction methods in the form of responding to the users' need for information become highly valued. Electronic shops and services offered by these websites are increasing. Agents are highly used in English electronic sales websites, but so far, there has been no research on designing an agent for extracting information in Persian for electronic shops. This study presented a model which receives the user's required conditions and browses for the goods in accordance with user's need. Using this model leads to increasing browsing speed and providing the goods related to the user's need.

Acknowledgement

This paper is based on a research plan titled as designing personalized assistant agent for information extraction of the user's required data that is being proceeded affiliated to Falavarjan Islamic Azad University.

References

[1] Rimmel G, Clement M, Runte M. Intelligent Software Agents Implication For Marketing In Ecommerce. *Springer Verla*.2000: 19- 33.
[2] Dzeng RJ, Chun Lin Y. Intelligent agents for supporting construction procurement negotiation. *Computer Law & Security Report.* 2004; 20(1): 20-27.
[3] Pivk A, Gams M. E-commerce Intelligent Agents. Communications of the ACM.1999; 42(3): 79-80.
[4] Kowalczyk R, Ulieru M, Unland R. Integrating Mobile and Intelligent Agents in Advanced e-Commerce. 2013; A Survey, Available online http://www.old.netobjectdays.org
[5] Kowalczyk R, Ulieru M, Unland R. Integrating Mobile and Intelligent Agents in Advanced e-Commerce: A Survey, Available online http://www.old.netobjectdays.org, 2000.
[6] Fard F, H Far, BH. *A method for detecting agents that will not cause emergent behavior in agent based systems-A case study in agent based auction systems*. 13th International Conference on Information Reuse and Integration (IRI). 2012.
[7] Gasparovic, BMezei I. *Auction aggregation protocols for agent-based task assignment in multi-hop wireless sensor and robot networks*. International Conference on Advanced Intelligent Mechatronics (AIM).July 2011 .
[8] Xuwang Liu,Dingwei Wang. *The behavior analysis of grouped multi-attribute auction based on multi-agent.* 24th Chinese Control and Decision Conference (CCDC). 2012.

[9] Wan C, Cheung R. *An auction agent architecture for mobile commerce.* 2nd International Conference on Education Technology and Computer (ICETC). 2010.
[10] Sandu L, Sbora R, Ilie S, Badica C. *Scalable distributed agent-based English auction server*. 15th International Conference on System Theory, Control, and Computing (ICSTCC). 2011.
[11] Li B, Ma Y. *An Auction-based Negotiation Model in Intelligent Multi-agent System*. International Conference on Neural Networks and Brain. 2005.
[12] Huang J,Youliu D, Yang B. *Online autonomous auction model based on agent.* Proceedings of International Conference on Machine Learning and Cybernetics. 2004.
[13] Akkaya Y, Badur B, Darcan O. *A study on internet auctions using agent based modeling approach*. International Conference on Management of Engineering & Technology. 2009.
[14] Qian D. *The principal-agent relationships in group buying auction.* 2nd International Conference on Management Science and Electronic Commerce (AIMSEC). 2011.
[15] Cheung R, Wan C, Cheng C. *An Active Learning System for Auction Agent Programming with Competitions.* 9th International Conference on Computer and Information Science (ICIS). 2010
[16] Yue C ,Mabu S, Chen Y, Wang Y. Agent bidding strategy of multiple round English Auction based on genetic network programming.ICCAS-SICE. 2009.
[17] Mesbah S, Taghiyareh F. *A new sequential classification to assist Ad auction agent in making decisions*. 5th International Symposium on Telecommunications (IST). 2010.
[18] Chan H, Ho ISK, Lee R*. Design and implementation of a mobile agent-based auction system*. IEEE Pacific Rim Conference on Computers and signal Processing, ACRIM. 2001.
[19] Xiong G, Okuma S, Fujita H. *Multi-agent based experiments on uniform price and pay-as-bid electricity auction markets*. Proceedings of the International Conference on Restructuring and Power Electric Utility Deregulation Technologies. 2004.
[20] Yunfei Gong. *Automatic web page segmentation and information extraction using conditional random fields.* in 16th International Conference on Computer Supported Cooperative Work in Design.2012.
[21] Jellouli I. an ontology-based approach for web information extraction. Colloquium in Information Science and Technology (CIST).2010.
[22] Hao Han. *A Method for Integration of Web Applications Based on Information Extraction.* Eighth International Conference on Web Engineering.2008.
[23] Liu Li. *Web Information Extraction Algorithm Based on Ontology and DOM Tree.* International Conference on Computational Intelligence and Software Engineering.2010.
[24] Xinchao Han. *Research on Web information extraction based on spider algorithm and DOM thinking*. International Conference on Information Networking and Automation (ICINA).2010.
[25] Yi Wei-Guo ,Yan Ling-Wei, Liu Ya-Qing, Liu Zhi, *An ontology-based Web information extraction approach*. 2nd International Conference on Future Computer and Communication (ICFCC). 2010.
[26] G Prassas, K C Pramataris, O Papaemmanouil, G J Doukidis. A Recommender System for Online Shopping Based on Past Customer Behaviour. Expert Systems with Applications 27 (2004) 365–377. Available online: http://www.elsevier.com/locate/eswa.
[27] M Vetter1, S Pitsch. T*owards a Flexible Trading Process over the Internet Agent- Oriented Software Engineering.* Second International Workshop, AOSE, 2001
[28] M Brian Blak. Towards the Use of Agent Technology for B2B Electronic Commerce .available online http://daruma.georgetown.edu, JUN-02