

## Toward a user-oriented recommendation system for real estate websites

Xiaofang Yuan, Ji-Hyun Lee\*, Sun-Joong Kim, Yoon-Hyun Kim

Graduate School of Culture Technology, KAIST, Daejeon 305-701, Republic of Korea

### ARTICLE INFO

#### Article history:

Received 30 July 2012

Received in revised form

16 August 2012

Accepted 17 August 2012

Available online 11 September 2012

#### Keywords:

Home buyer

Real estate website

Housing search behavior

Case-based recommendation system

Ontology

### ABSTRACT

The Internet has become a significant transaction platform for the real estate industry. However, use of the Internet does not benefit homebuyers in terms of search time, flexibility, and intuitive results. While it does encourage buyers to search more intensively, and discover and visit more properties, it also wastes more time and energy. To improve the efficiency of real estate searches, we developed an online homebuyer's search program, based on an investigation of search behaviors, and implement a user-oriented recommendation system for real estate websites via a combination of case-based reasoning and an ontological structure. An ontological structure is employed to improve information management efficiency while case-based reasoning improves recommendation accuracy. A user test demonstrates the effectiveness of the proposed system and validates the findings of this study. The limitations of the current study are also discussed for future research and applications.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

With the prevalence of the Internet, more Internet-enabled services are becoming available. The American National Association of Realtors (NAR) On-line Technology Survey, conducted in 2011 ([34] National Association of REALTORS® Profile of Home Buyers and Sellers), indicated that when searching for a home, 88% of homebuyers chose the Internet as an information source, an increase of 14% since 2010. Real estate websites have thus been used more frequently in the residential real estate industry. Nevertheless, the median time that homebuyers spent on a home search was 12 weeks, the same as in 2009 ([35] National Association of REALTORS® Profile of Home Buyers and Sellers). This implies that the use of real estate websites did not improve homebuyer's search efficiency [11,27]. In Korea, the real estate market territory of web-based real estate search tools has expanded due to advantages of

information management and ease of provision to real estate customers [22,7]. In particular, technological innovations have resulted in new applications that allow homebuyers to select affordable price ranges, take virtual tours, obtain information about the neighborhood and surrounding environment, access comparative data of different districts, and view property values over time.

While there are services that can connect multiple real estate agencies with customers and provide varied economic information for real estate, such services only provide limited knowledge for customers. For example, the existing linear search mechanism and housing search process services restrict the ability to run a multi-attribute housing search. As with other searches, intricately connected factors are used to make housing decisions [27,38]. In Korea, web-based real estate search services have been proactively developed due to the nation's well-constructed Internet infrastructure [22,7]. Although there are many indications for the need of a multi-attribute search system, current Korean housing search services still only allow single-attribute searches [7].

\* Corresponding author. Tel.: +82 42 350 2919; fax: +82 42 350 2910.  
E-mail address: [jihyunlee@kaist.ac.kr](mailto:jihyunlee@kaist.ac.kr) (J.-H. Lee).

Ontology allows a semantic evolution of search engines, providing a systemic way to implement multi-attribute housing searches [15] among intricately connected sub-information, and describing common features shared by multiple housing units. The semantic relationships between common information will increase the efficiency and accuracy of information and result in a housing recommendation system. A multi-attribute recommendation system can expand customer's knowledge via similar case-based experiences of other users [7].

Nevertheless, there are unresolved issues causing inefficiency and lack of knowledge support for customer decision making on conventional real estate websites, including the following.

- Home purchase is a multi-facted decision, but current search services are single-faceted, thus inhibiting customers when comparing properties.
- Individually constructed housing related information cannot deliver semantic meanings between common information segments. A linearly structured information database has particularly limited information storage efficiency. Despite that home assessment involves intricately connected information, current services do not express semantic relationships between various factors.

To improve the efficiency and affordability of an online housing search, a user-oriented recommendation system for real estate websites, using ontology and case-based reasoning (CBR), was designed. This paper, as the first half of our whole research, focuses on user requirements and search behaviors. This problem addresses not only the semantic construction of housing unit information but also all sub-mechanism designs for constructing a recommendation system. Specifically, this study deals with practical problems that occur with the design of a multi-attribute housing unit search system and the construction of semantic meaning in order to develop a CBR-adapted recommendation system. To do this, the methodology was divided into four parts. First, the functions and search procedures of current web-based housing search engines were analyzed. Second, through a questionnaire and direct observation, we investigated user search behaviors to extrapolate user requirements. Third, the semantic relationships between information segments were represented in an ontological structure from the knowledge collected in the user study and real estate domains. Finally, the case representation was clarified and case indices were specified to calculate similarity between homebuyer's queries and cases.

The remainder of the paper is organized as follows. We begin in Section 2 by looking at previous work on understanding web-based housing search behavior and case-based recommendation systems. Next, in Section 3, the methodology is explained in detail and our model of homebuyers' online search behavior is described. In Section 4, we implement our system prototype. We conclude, in Section 5, with a discussion of the potential implications of our research and possible directions of future work.

## 2. Related work

### 2.1. Internet housing searches

#### 2.1.1. Web-based housing search systems

Online information search behavior has been studied mostly in the context of information retrieval. While some empirical studies have examined online search behavior in real estate markets, but as the successive development of research on markets, they predominantly focused on intermediary and seller searches. In particular, the emphasis has always been on selling price and the Time on Market (TOM) [38].

Bond et al. [3] investigated the listing information of existing real estate brokerage websites (Fig. 1), including information on geographic region, price, square footage, atypical features (e.g., fireplace, swimming pool, gas versus electric), financial factors (property tax information, insurance, association fees), information on recent comparable sales, information on the surrounding area (schools, shopping, recreation), types of photos available (house exterior and interior), and other miscellaneous factors (mortgage calculator, prequalifying, rent versus purchase). These categories represent most of the information on real estate websites. Based on this, we tabulated the necessary information for a real estate website, combining analyses of the three most widely used real estate websites in Korea.

Leonard et al. [27] first explicitly examined intermediation in terms of the Internet's effects on a buyer's housing search. They proved that broker intermediation reduced overall buyer search time and increased search intensity; in contrast, Internet intermediation improved buyers' search intensity, but did not reduce search time. Ford et al. [18] compared the selling price and TOM of listed properties on the Internet, while listing the same properties on multiple listing services. Their study indicated that Internet listings remained on the market about 11% longer than houses not listed on the Internet, but sold at higher prices.

To date, however, no research has explicitly uncovered the housing buyer search process and preferences on the Internet. Therefore, this paper investigates real estate websites as a search tool and their impact on all search dimensions. To accomplish this, we generated an online buyer search model.

#### 2.1.2. Users' internet search behavior

Since the Internet has become a primary source of information for many people [21,26], much attention has already been devoted in the literature towards understanding users' Internet search behavior. A number of independent variables that influence user search behavior have been proposed in the IR domain [24]. For example, Hsieh-Yee [20] reviewed studies conducted between 1995 and 2000 on web search behaviors, reporting that many studies have investigated the effects of certain factors on search behavior, including information organization and presentation, search task type, web experience, cognitive abilities, and affective states. Liaw and Huang [28] claimed that individual experience and motivation, search

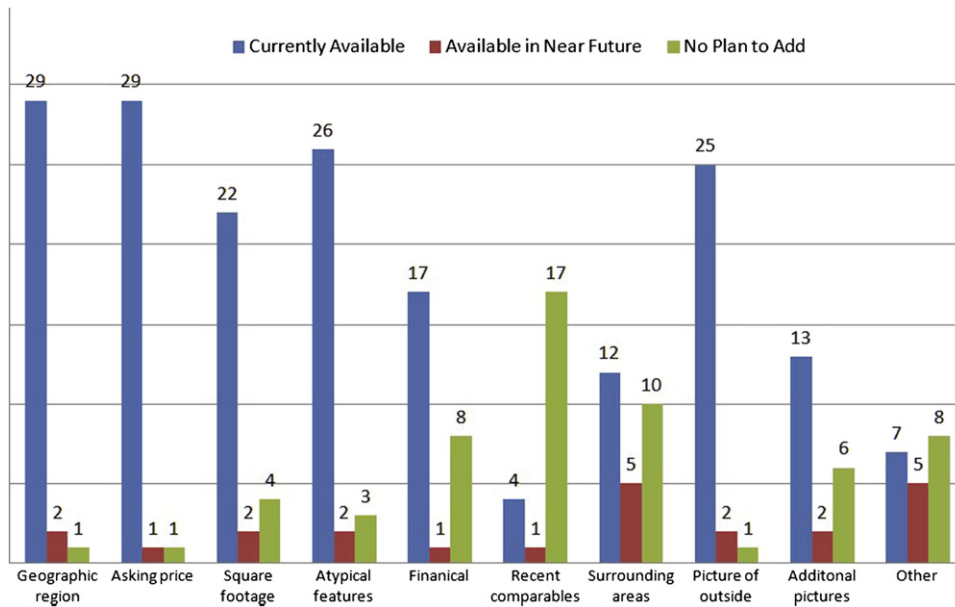


Fig. 1. Listing information of existing real estate brokerage websites [3].

engine quality, and user perceptions of technology acceptance are all factors affecting individual desire to use a search engine. Jansen and Spink [21] investigated the interaction between users and search engines through a comparative study of nine major search engines over a seven-year period.

Choo et al. [8] developed an integrated behavioral model of web-based information seeking founded on modes of browsing and searching, differentiated by information needs and search activity. This behavioral framework contains the scanning modes, identified by Aguilar [1], Weick and Daft [36], and six categories of information seeking behaviors, which were identified by Ellis [12] and Ellis et al. [13] Ellis and Haugan [14].

## 2.2. Case-based recommendation system

Recommendation systems were originally described as “people providing recommendations as inputs, which the system then aggregates and directs to appropriate recipients” [32]. Subsequent research expanded the connotation into all the systems which can provide users with customized recommendation and personalized search method in an almost infinite database of possible options [6]. Recommendation systems combine ideas from “information retrieval and filtering, user modeling, machine learning, and human-computer interaction” [4]. Recommendation techniques have a number of classifications [32,33]. The most commonly used recommendation techniques are collaborative, demographic, knowledge-based, and case-based. Collaborative and case-based approaches are the two main techniques employed by recommendation systems [4].

Case-based reasoning (CBR) is a problem solving methodology that addresses a new problem by retrieving similar solved cases and reusing that information for solving current problems. CBR contains four main steps:

retrieve, reuse, adapt and retain [23]. CBR has played an essential role in the development of an important recommendation system known as “case-based recommendation system”. The common usage scenario for case-based recommendation systems is that a user is searching for something. The user inputs some requirements about the item being searched for and the system searches the case-base for items that match the requirements. The similarity calculation of items or problem descriptions drives case-base retrieval. A set of cases is retrieved from the case-base and shown to the user in a sequence organized by similarity [29].

Lorenzi and Ricci [29] proposed a general framework for case-based reasoning recommendation systems (CBR-RSs), which include the classical five steps of CBR problem solving cycles, plus an additional “iterate” step. Some CBR-RSs analyze using the framework to find common features. The framework proposed here can help describe to what extent a recommendation system exploits the CBR cycle.

In CBR-RSs, item descriptions are vital to generate a set of recommendations by retrieving items whose descriptions match the user’s query. In most CBR-RSs, case base model items to be recommended and sets of suggested items are retrieved by searching for items similar to those partially described by the user [5]. In these approaches, a case and an item are considered identical, and item features represent the main components of the case.

## 2.3. Combination of ontology and case-based reasoning

CBR systems can more intelligently infer fittings for a user’s needs, but it is difficult to clearly specify a case. In past CBR systems, cases are restricted to the database table. In other words, they cannot describe relationships among several variables organizing the case; it is hard to

provide flexible and contextually correct results [25]. However, describing the relationships between cases and each case element is necessary to deliver an accurate result. There is, accordingly, a need for adaptation of semantic meaning in system databases.

Recently, to create semantic association among variables to address these problems, research on the use of ontology for CBRs has been pursued [10,31]. Ontology can enable quality work automation including information search, interpretation, and integration. It allows people and computers to easily understand information by providing meaning between diverse, distributed information throughout the Web [2]. Therefore, an ontology offers advantages in systematically organizing cases, and thus building a knowledge model [10,19,31].

In research that has focused on ontological concepts for CBR system design, the usefulness of domain knowledge construction was mainly considered. In particular, the semantic structure of an ontology can deliver the relationships between vocabularies at the each phase: querying, retrieving, similarity measuring, adapting, and learning [31]. The semantic structure of a CBR system is usually constructed by domain knowledge that represents the relationships between variables. This is because a CBR system is domain-specific and requires circumstantially defined relationships.

A semantically implemented CBR concept can adopt Methontology, a structural methodology for ontological knowledge construction. As a systemic solution for ontological construction, Methontology has been introduced in the domain of ontological engineering [16]. Methontology has specific tasks to construct ontological knowledge models as a methodology. Corcho defined the construction tasks according to 11 steps, from building a glossary of terms to describing instances [9]. All components of an ontology can be generated by a methodology founded on Methontology; the essential components are concepts, attributes, relationships, constants, formal axioms, rules, and instances [9].

### 3. Methodology

#### 3.1. Functional analysis of existing search services

Housing unit search services have a search engine structure that is approachable on the Web. Like a search engine, the functions that are provided represent the range of user need allowance. To analyze the limitations of existing search engines, the functions they provide need to be determined. Further, the hierarchical procedure that users follow should be extracted and described.

In Korea, three major real estate search engines are broadly used: *Budongsan 114*, *Speed Bank*, and *Doctor Apart*. Each service provides different processes for intended housing unit searches. Table 1 describes the processes and results of existing services.

Linear processes resulted in an inflexible search experience, which is the most problematic issue for homebuyers during their search. If a buyer decides to change one attribute, he or she must return to the initial page and restart the entire search process.

Additionally, the services only support specific location-based searches. The linear search process, as described in Fig. 2, only prints out the results by location and price input constraints. This restricts users, not allowing them to broaden their location knowledge to unknown sites. Usually the location of appropriate housing units is decided by the background preferences of each user. Even though user preferences act as search constraints, current search engines only support pre-narrowed searches.



Fig. 2. Summarized common search procedure of three primary Korean search engines.

**Table 1**  
Search process and results of existing real estate websites.

Engines	Searching process	Result
<i>Budongsan 114</i>	1. Choose location: province (Do)→city(Si)→district(Cu)→town(Dong) 2. Choose area property (optional) 3. Choose price (optional)	Price fluctuations + Housing unit's information (text)
<i>Speed bank</i>	1. Choose location: province→city→district; or choose the Metro route 2. Choose area property (optional) 3. Choose price (optional)	Price fluctuations + Housing unit's information (text) + Environmental facilities (visualized on map)
<i>Doctor Apart</i>	1. Choose location: province→city→district→town 2. Choose area property (optional) 3. Choose price (optional)	Price fluctuations + Housing unit's information (text) + Environmental facilities
	Choose by themes: Low price, Enrolled date, Metro Line, Schools & Good floors	

In general, the search process can be summarized as follows; Steps 2 and 3 may shift according to different systems (Fig. 2).

### 3.2. User search behavior study

#### 3.2.1. Subjects of the user study

To understand user search behavior, 30 participants (11 female and 19 male) were recruited for our survey via an advertisement on notice boards from a South Korean university. Of the participants, 22 were graduate students, 5 were researchers, and 3 were Ph.D. students. All of the participants had online housing search experience, including purchase (19), rental (9), and investment (2). Their ages are from 22–46 ( $SD=5.1$ ), the minority (46%) of home buyers in the demographic survey of *National Association of Realtor* in 2011, while the mean age is 28.6.

#### 3.2.2. Data collection and analysis

Before the search process began, we used similar instructions suggested by Ericson and Simon (1993) to explain the procedure to participants. Following the instructions, the participants practiced both search and “think-aloud” (response) procedures for this study.

Each of the participants described his/her housing search. A scenario-based questionnaire, with short instructions, was then handed out (Fig. 3). After reading the instructions and completing the questionnaire, each participant was asked to perform an online housing search on a computer with a fast Internet connection. During the search, participants were asked to think-aloud and explain their responses. Their search behavior was recorded by a video camera, and the observer/interviewer noted their searching sequence in steps, as well as their complaints while searching. Meanwhile, cookies recorded all the pages they visited.

The data were collected by one-to-one interview, and each participant took at least 1 h for the survey. This task ensured that we could record the complete online search process as well as reveal most of the problems homebuyers would meet during an online housing search. At the end of the experiment, the participants were briefly

interviewed again about their experiences during the search session and asked to confirm their concerns and inconveniences, while an observer simultaneously listed their responses.

#### 3.2.3. Personas and requirements

From our data, we derived three representative scenarios that imply the search requirements and constraints of each defined persona.

Three representative personas are summarized as follows:

- A working couple with two children, one of the children is school-age.
- A working couple without children, enjoying urban “high-life”: fashion, entertainment, and convenience.
- Single, needs a small but well-constructed home, prefers community with amenities.
- The prominent commonality of the three personas is that they want to be within a certain distance to a desired location. Hence, the core concept of the proposed system is that it is location and distance-based. Further details are provided in Chapter 4.

The requirements were broken into three categories: Location, Price, and Housing Unit Property (Table 2).

#### 3.2.4. Model of homebuyers' online search behavior

The integrated behavioral model of web searching developed by Choo et al. [8] identified four modes of online information gathering: undirected viewing, conditioned viewing, and informal and formal searches; the six steps of information processes were: starting, chaining, browsing, differentiating, monitoring, and extracting. Because an online housing property search is a kind of information search, we followed Choo's behavioral model, combined it with our user research, and built a homebuyers' online search behavior model, as shown in Fig. 4.

Homebuyers start from their favorite real estate website, select from the queries-chaining, and check the search results. They then differentiate the results, checking the details of some satisfactory results. Finally, they extract the best ones and contact the agent/sellers.

**Housing Search Task**

Your considerations and behaviors that you experience in a web-based housing search process are gathered in this interview. You may explain the considerations – constraints and preferences – with a format of scenario. And the behavior will be collected by the housing search task that is based on your own scenario.

First, describe the constraints and preferences that you consider during the housing search process. You may imagine a scenario for housing search or describe the experience that you had before. Clear expression is recommendable. For example, if you would like to find a house which is located in a 'suburban area' then please specifically explain the constraints – transportation, environment, etc. – of the considerable places to define the appropriate alternatives.

Second, according to the scenario please search the area and housing unit that are appropriate for your constraints and preferences. Use the search engine that is indicated. During the search, you should verbally express the each phase of search, including the goals of each phase and the evaluations for each response.

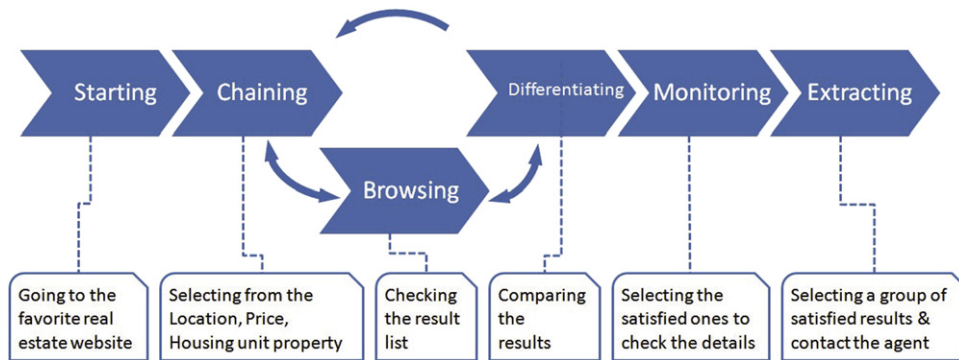
Remember, you have 1 hour to describe a scenario and search a house. The interviewee will announce you the elapse of time. A video recording and interview will be used to gather information.

Fig. 3. Instructions for the housing search task.



**Table 2**  
Representative requirements and constraints.

User requirements	Working couple with children	Working couple without child	Single
Location	Convenient to working place Walk to the elemental school and kindergarten within 20 min Walk to the bus stop within 10 min, prefer subway	Prefer the house nearby the working place Prefer living in the downtown Prefer the house near the shopping places such as supermarket and mall Prefer the houses nearby the bus and subway station	Prefer the house nearby the working place Prefer the house near the shopping places such as supermarket and mall Prefer the house nearby the bus and subway station
Housing unit property	Prefer three bedrooms (parent, child 1, child2) Prefer one living room and at least one bedroom are on the sunny side Prefer 1 or 2 bathrooms with toilet and sink, only one shower booth is acceptable Storage space, prefer wall embedded storage Parent's room and children's rooms are adjacent Prefer reconstruct able balcony as the children grow up Prefer well-constructed noise prevention	Prefer two rooms, the space is not important Storage space, prefer wall embedded storage Prefer the additional space that allows natural drying (e.g. balcony) Prefer higher stories Prefer good scenery viewed house Prefer the building that has enough parking lots	Prefer one room (or two rooms) Prefer built-in residential house Prefer the building that has satisfied the safety issue Prefer well-constructed ventilation system Prefer well-constructed noise prevention
Price	Depend on customer's economic status	Depend on customer's economic status	Public imposts included monthly rent Depend on customer's economic status



**Fig. 4.** Model of homebuyers online search behavior.

According to the cookie record and observation, we found that, on average, homebuyers initially selected a certain location, price, and housing unit property queries, but when they saw the results, they returned to the homepage and adapted one or all of the attributes and re-checked the results. In this sense, the search process of selecting attributes and re-checking results is an iterative process. Homebuyers might change one of the selected items during browsing and differentiation of results

### 3.3. Ontological information construction

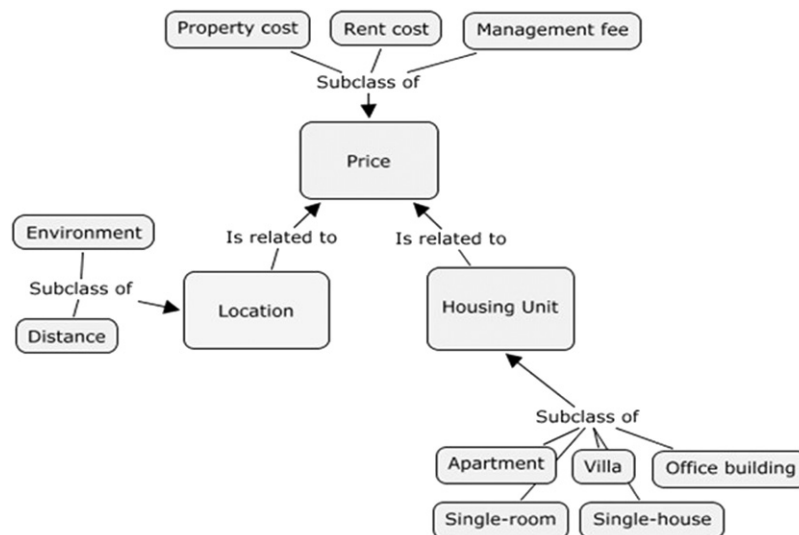
The customer requirements and constraints that were revealed via a scenario analysis were used to construct an ontology framework by organizing keywords. Using this framework, we composed a concept dictionary and concept classification tree (Table 3 and Fig. 5). In this study, Methontology was used to construct an ontological

structure for storing and reusing domain information. We used the ontology construction method of Corcho (2005) to define concepts, attributes, relationships, and constants. In the adaptation of the CBR concept, formal axioms, rules, and instances were defined by a similarity measurement strategy. Highly required domain specific knowledge – real estate domain knowledge – created difficulties when creating a universal methodology. Although methodology has been developed for adapting information semantic meanings, the domain specified strategies for the definitions of formal axioms, rules, and instances were used to imply domain knowledge in this research.

The concepts were classified into three main clusters in a concept dictionary: location, housing unit property, and price; each first level concept has its own sublevel (Table 3). The first resource cluster, “location,” includes the distance to an address and a location’s environmental

**Table 3**  
Concept analysis driven classification and its attributes.

Concepts			Attributes
First Level	Second level	Third Level	
Location	Distance	By bus By subway By taxi By private car On foot	10 minutes radius, 30 minutes radius, 1 hour radius, More than 1 hour radius
	Environment	Convenience environment Education Environment Transportation Facility Security Size (m <sup>2</sup> )	Government office, Mart, Hospital, Cultural Infrastructure, Park, etc. University, High school, Middle school, Primary school, Kindergarten Bus station, Subway, Railway station, Airport Playground, Resident center, Parking, Fitness center, Guard, CCTV, low crime rate
Housing unit property	Apartment		0-65, 65-100, 100-130, 130-165, more than 165
	Single-room Villa Single-house Office building	Bathroom Bedroom Balcony Living room Structure Orientation Heating method Ventilation Completion duration	1, 2, 3 or more 1, 2, 3 or more 1, 2, more than 2 1, 2 or more Multi-storey, Mid-rise, High-rise, Duplex South, South-east, South-west, East, East-south Central heating, Individual heating Natural ventilation, Artificial ventilation Within 1 year, Within 5 years, Within 10 years, More than 10 years
Price	Property cost		Ranges according to the market quotation
	Rent cost	Deposit rental Monthly rental	Ranges according to the market quotation
	Management fee		Ranges according to the market quotation



**Fig. 5.** Class hierarchy of upper ontology.

information. The distance to an address means the radial range of the places surrounding the destination within a certain time. The address is the destination that a user selects on the map; it can be either a building or a district. The second resource cluster, “Housing unit property,” consists of five classes (type of housing) and ten properties. The third cluster includes three classes of cost in buying/renting a housing unit; the rent has two subclasses: monthly fee rental and deposit-only rental, a special rental method widely used in South Korea. To do

this, words related to the real estate domain knowledge were classified by a concept taxonomy. The concepts that are used in housing searches were analyzed by using user scenarios gathered in our user study.

The concept attributes were defined by the information structures of common housing search systems for compatibility. For example, the “environment” concept has attributes of station, university, high school, cultural infra, working place, green area, etc. The binary relationships containing tree structures are represented in Fig. 5

and concept relationships follow the domain knowledge of real estate. The constants were defined in detail.

Our system ontology is divided into a common upper ontology for general concepts of real estate and specific ontologies that apply to different sub-classes. The upper ontology defines the basic concepts of location, housing unit, and price, and consists of ten classes, as shown in Fig. 5

The details of these basic concepts are defined in specific ontologies that vary from one sub-class to another. For example, “Location” has the sub-classes of distance and environment. We have defined two specified ontologies for each, as shown in Fig. 6.

### 3.4. Case representation and similarity measurement

A CBR-RS is composed of a base of cases, a retrieval module, and a recommendation module. The system receives a problem description, searches the case base, assesses the similarity of the cases and a new problem description, and then selects the best matches. The case base contains problem descriptions and solutions [30]. A problem description is the wish list that is chosen by a user when searching for items. Based on our investigation of user searching behavior and processes, the problem description in the developed system includes desired location, desired price, and desired housing unit property (Fig. 7).

- Location: users prefer housing a certain distance from a desired location.
- House unit property: what type of housing, how many baths or bedrooms, preferred orientation, etc.
- Price: the expected price of the housing and management.

After a user confirms his/her selection, an item wish list is produced. Once the user's wish list is produced, a similarity measurement of the queried cases and the problem solution database proceed in order to find the most consistent results for recommendation.

The location information of each apartment, including the distance and environment, is built as a search result

ontology according to the collected data. Two examples are shown below (Fig. 8).

The housing unit property information, including the number of rooms, the size, and the layout, is represented by the SEED Layout program [17]. SEED Layout was developed by the School of Architecture of Carnegie Mellon University in 1995 to assist designers in the schematic phase of layout design to deal with the overall organization and configuration of a planned building. We employed the SEED Layout program for two reasons: first, users can describe their specific expectation of the layout accurately and comprehensively; and second, the existing online real estate service data are only provided as a housing layout picture, and lack specific data such as size, adjacency, orientation, etc. Therefore, SEED Layout can be used to represent the cases via a numerical similarity measurement.

Users can access the SEED Layout program through our system and draw their expected layout, then input it into the system to search for matching houses. Query input and apartment layout cases are stored in the database via the SEED Layout file, “.loos.” Meanwhile, the text description of the layout, which is the evaluation window of a SEED Layout file, shows the layout numeric information and structural constraints that the SEED Layout has established as domain knowledge (Fig. 9).

We built a case base using more than 200 sets of apartments in Yuseong-gu, Daejeon, South Korea and more than 300 environment cases, including public office, parks, cultural infrastructure, schools, etc.

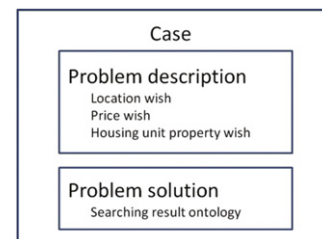


Fig. 7. Case representation.

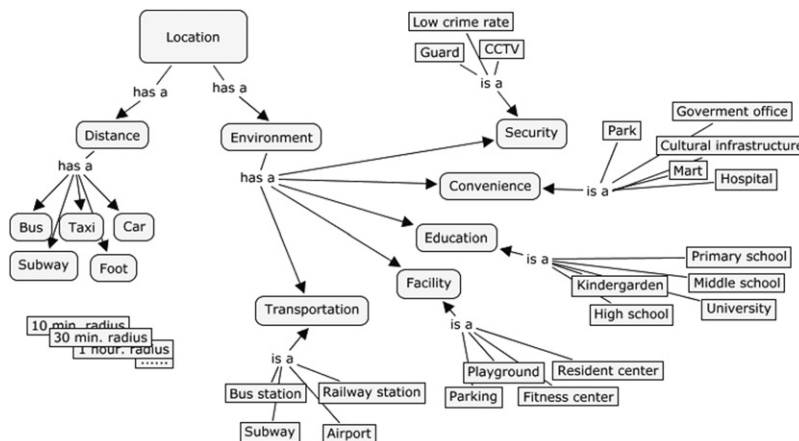


Fig. 6. A partial definition of a specific ontology for “Location.”.



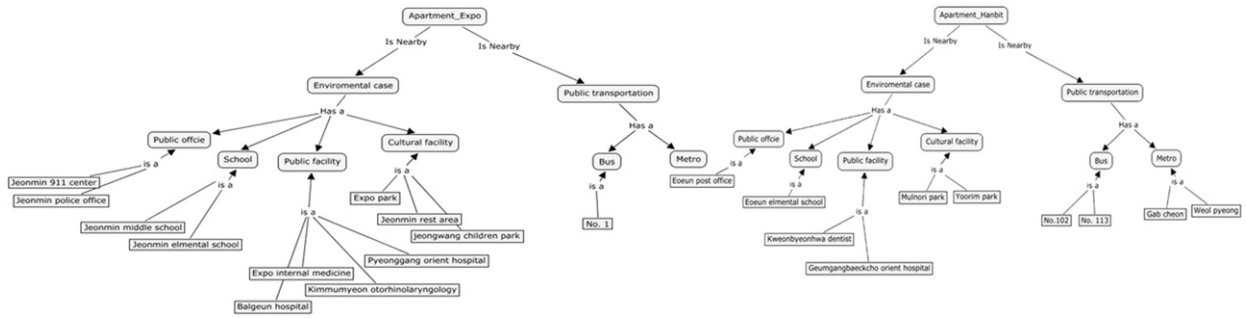


Fig. 8. Two example search results.

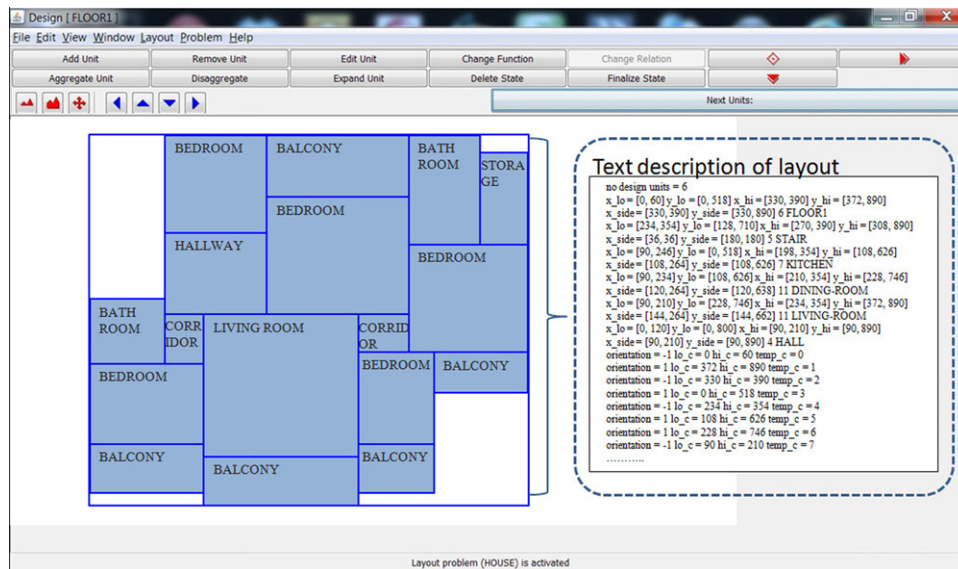


Fig. 9. Example SEED Layout "loos" file data and text description.

## 4. Implementation

### 4.1. System architecture

A web-based real estate search/recommendation system was implemented to improve the housing search efficiency. Our system consists of three parts: operation, data acquisition, and data processing (Fig. 10).

First, users input their preference information into the system through a web-based interface. Second, the query analyzer deals with the input information, searches the database, and displays the related results through the mash-up service module. Third, as the users narrow down their requirements, the similarity measurement module uses semantic and numeric measurements to compare the user's query and database cases in order to provide recommendations. Finally, the results are shown on the user interface.

The system was developed from PHP and Java script programming language and a MySQL Database. The system has been tested in IE 8.0 and Google Chrome.

### 4.2. GUI and manipulation

#### 4.2.1. Map-based interface

Since location is the most important issue for homebuyers, we built the system GUI as a map-based interface. The default scale of the map is 1:1000, which is the urban scale for site or location plans (Farrelly, 2007). Our GUI is developed based on the Daum open API mash-up service, and consists of three searches: location, price, and housing unit property. The map is displayed on the left, and the price and housing unit property query are on the right (Fig. 11). Once the user selects a location on the map and inputs a query, the homes that fulfill the criteria set by the user's current query are shown as dots on the map.

#### 4.2.2. Scenario and manipulation

Prior studies of real estate industries and our user investigation verified that location is the prominent factor for homebuyers during a housing search. We believe that a map is an appropriate way to visualize location information. Furthermore, direct manipulation is an easy to

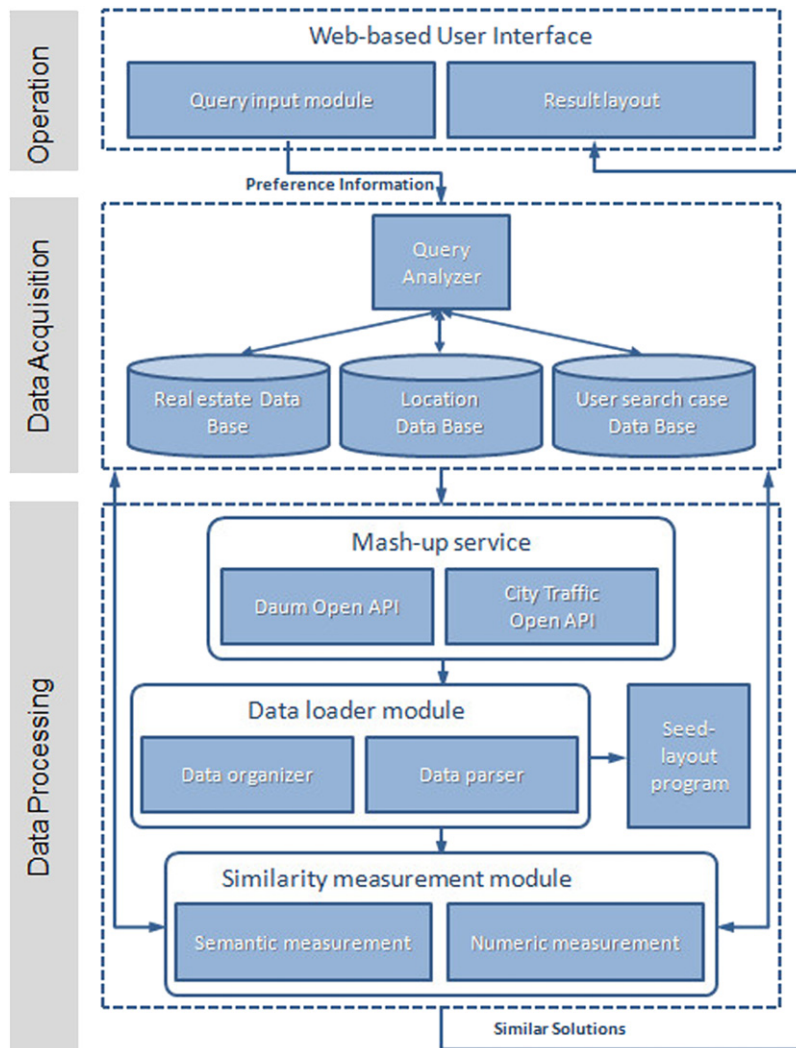


Fig. 10. System architecture.

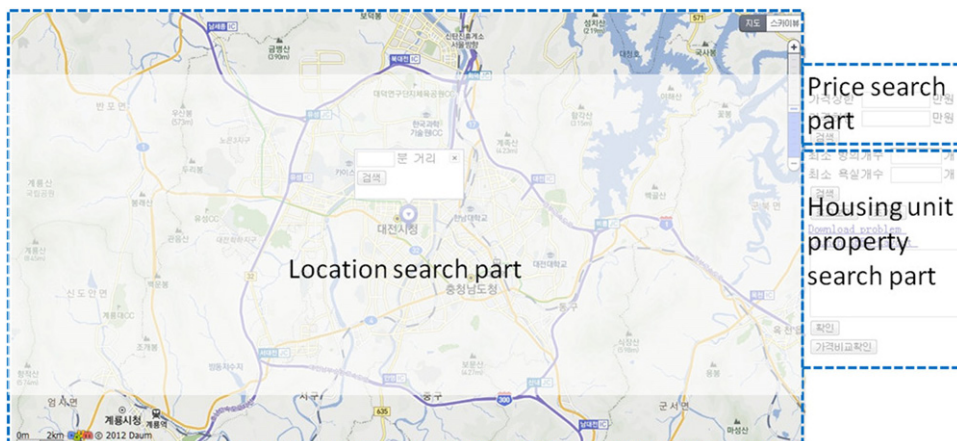


Fig. 11. GUI prototype of the proposed system.

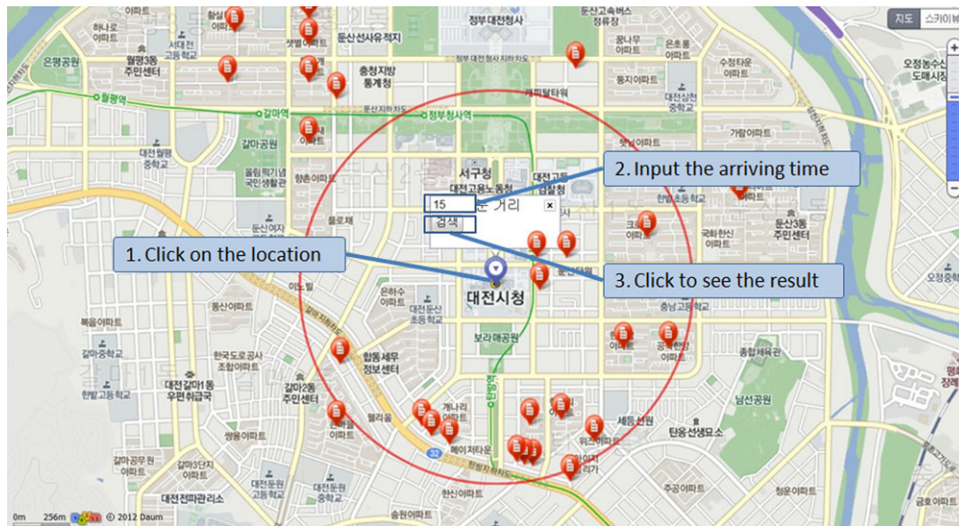
use, quick, and powerful query method for database and information retrieval [37]. Hence, through a combination of a map-based interface and direct manipulation, our system provides a visualization of both the query formulation and corresponding results.

We can use a hypothetical scenario to explain our system. Kim is a new public officer hired by the Daejeon City government, and he will use the proposed system to search for a home. He points out a location on the map where he prefers to live. Next, he selects the estimated distance (15 min) to a location in a popup window, in this case, Daejeon City Hall. When he clicks on the search button, all the homes within 15 min of City Hall are displayed on the map (Fig. 12). We used the traffic open API mash-up service to calculate similarities of different forms of transportation. We also contacted the Daejeon Traffic Policy Bureau and obtained authorization for Daejeon City's open API public transportation data. In this

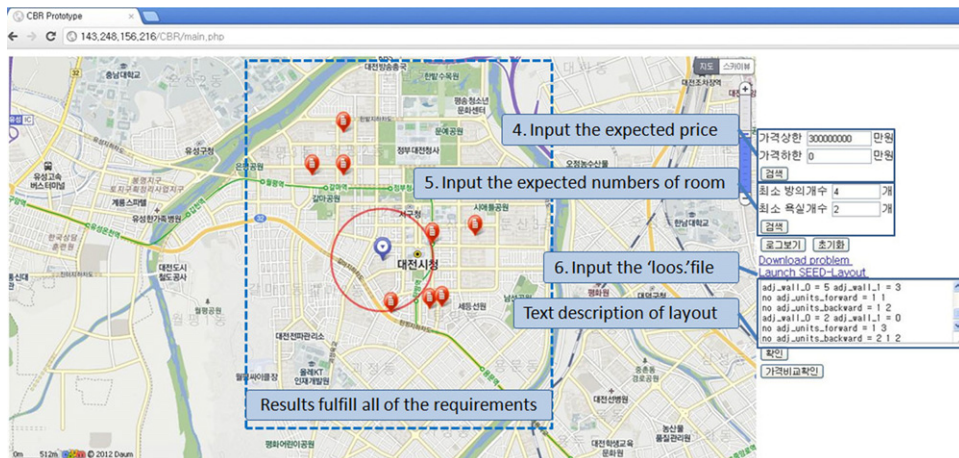
manner, we can find all homes possessing similar accessibility to a certain location by all kinds of regular transportation, including walking, bus, subway, taxi, or private car. The red circle indicates the walking area range, and red balloons are home icons.

Kim wants to buy an apartment under 300 million Korean won, and thus he inputs the maximum price in the price box. As he clicks on the search button, he can see homes are eliminated according to his price constraint. Finally, he requires four bedrooms and two bathrooms in the apartment, and he inputs this requirement in the housing unit property part. The remaining results fulfill all of Kim's desires, and are indicated by red balloons on the map.

Furthermore, if Kim has some specific requirements for the housing unit property, then he can easily launch the SEED Layout program, which is embedded in our system. He can then make his expected layout in SEED Layout, and



**Fig. 12.** All the homes within 15 min of Daejeon City Hall. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)



**Fig. 13.** Close-up of the system with all homes within a 15 min radius of Daejeon City Hall, possessing four bedrooms and two bathrooms, costing less than or equal to ₩300 million, and containing the desired layout.



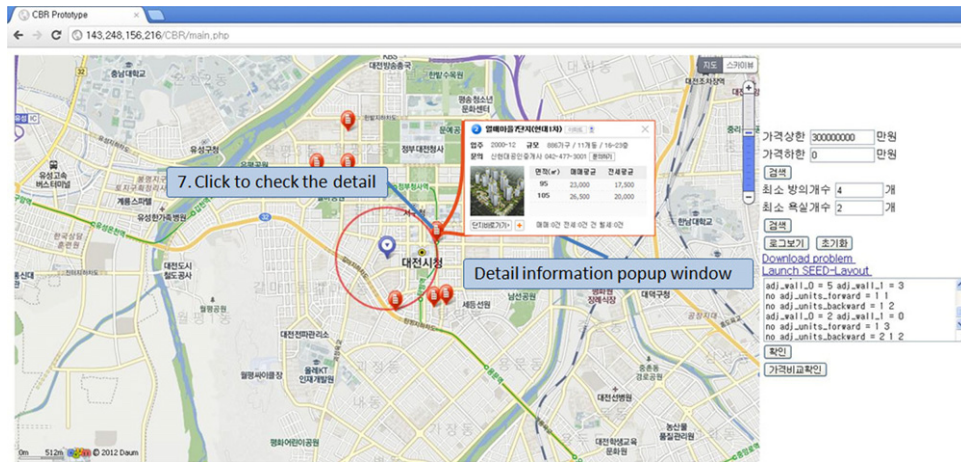


Fig. 14. Detailed information in a popup window.

**Table 4**  
User evaluation results of the GUI.

Questions	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1 Do you feel easy to use the interface?	0	0	2	12	16
2 Can you use it without training?	0	0	3	9	18
3 Do you accomplish the search more quickly?	0	0	0	8	22
4 Are you satisfied with the interface?	0	0	7	15	8
5 Are you satisfied with the search process?	0	0	1	6	23

input his “loos” file into the system to search for similar cases in the database. Recommended results are displayed on the map, and a corresponding text description will appear to the right of the interface (Fig. 13).

To keep user search activity consistent, a mouse click on any of the homes will bring up detailed information about that specific home in a popup window (Fig. 14), such as photos, orientation, heating method, etc.

In this way, once a user narrows his/her query, it is easy to change the district, and also see what is available if he/she is willing to pay more. This will encourage users to explore different homes instead of the current laborious task of reading figures.

#### 4.3. Evaluation of the GUI

In a follow-up study, the same participants repeated the online housing search. The user test was still a one-to-one in-depth interview, and five items were assessed on a five-point Likert scale, indicating agreement descriptions of each item. The results of the GUI evaluation are shown in Table 4.

Over 70% of the participants were satisfied with the GUI. In particular, the users believed that the searching process is optimal compared to services they had previously used.

### 5. Discussion

In this study, we developed a homebuyers' search behavior model and integrated it into a web-based real estate recommendation system. Existing web-based

housing services were analyzed to diagnose issues. The necessary knowledge used in home searches and purchase were gathered by user studies involving a scenario description and scenario-based housing unit search. Via user studies, knowledge attributes and concepts were defined. The semantic relationships of concepts and attributes were defined in an ontological structure, following the systemic methodology of Methontology. Finally, we integrated the housing recommendation system into a map-based user interface and direct manipulation operation panel.

- The designed search system supports multi-attribute searches. As mentioned above, current single-attribute based searches do not satisfy user decision-making needs. Derived from user studies, “location,” “environment,” and “price” were defined as primary concepts considered in a housing search. Providing these three concepts as search limits enhances the search environment and supports a multi-attribute search.
- Adapting the semantic meanings into a recommendation system better supports recommendation results and more efficient storage of domain knowledge. The semantic meaning of domain knowledge was adapted from a CBR-based recommendation system by using an ontological structure. The concepts and attributes of a housing search were gained through user studies to construct an ontological structure. Current housing search engines only support text-based information. This blocks other similar alternatives due to a lack of definitions. Hence, we employed semantic matching in order to provide more flexible alternatives for homebuyers.

This evaluation verifies that our system improves the efficiency and affordability of homebuyer online housing searches. Nevertheless, this study still has some limitations. First, homebuyers' decisions are affected by many factors not only the distance to a location but also other environmental factors such as security, amenities, convenience facilities, etc. We only adopted the similarity measurement of location accessibility, and only public transportation (subway and bus) data were accessible for the current research. The similarity measurement of other environmental factors should be included in future work, and more data will be required as our project continues. We collected environmental data, including public offices, cultural facilities, and markets, from several online real estate agencies (www.r114.com, www.drapt.com, www.bd119.com), thus limiting the scope of environment data. In addition, we set clusters of each sub-class somewhat arbitrarily. Although we discovered some requirement descriptions in the interviews, there have been few studies on mechanisms for determining the optimal number of categories of these clusters. Attempts to adjust the number of clusters and categories should be a focus of future research. In addition, with the use of a different database, we believe that the same search/recommendation system may be used for sellers and tenants.

## References

- [1] F.J. Aguilar, *Scanning the Business Environment*, Macmillan Co, New York, 1967.
- [2] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web – a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, *Scientific American* 284 (5) (2001) 34–43.
- [3] M.T. Bond, M.J. Seiler, V.L. Seiler, B. Blake, Uses of websites for effective real estate marketing, *Journal of Real Estate Portfolio Management* 6 (2) (2000) 203–210.
- [4] D. Bridge, M.H. Goker, L. McGinty, B. Smyth, Case-based recommender systems, *Knowledge Engineering Review* 20 (3) (2006) 315–320.
- [5] R. Burke, Knowledge-based recommender systems, in: A. Kent (Ed.), *Encyclopedia of Library and Information Science*, Marcel Dekker, New York, 2000.
- [6] R. Burke, Hybrid recommender systems: survey and experiments, *User Modeling and User-Adapted Interaction* 12 (4) (2002) 331–370.
- [7] J.-H. Cho, The development of a real estate multi-attribute integrated search system, *Journal of the Society for e-Business Studies* 14 (3) (2009) 15–37. in Korean.
- [8] C.W., Choo, B., Detlor D., Turnbull (1999). Information seeking on the web: an integrated model of browsing and searching, in: *Proceedings of the 62nd Annual Meeting of the American Society for Information Science (ASIS)*, October 31–November 4, Washington, D.C., USA.
- [9] O. Corcho, M. Fernandez-Lopez, A. Gomez-Perez, A. Lopez-Cima, Building legal ontologies with METHONTOLOGY and webODE, in: V.R. Benjamins, P. Casanovas, J. Breuker, A. Gangemi (Eds.), *Law and the Semantic Web*, Springer-Verlag, Berlin, 2005. in:.
- [10] B., Diaz-Agudo P.A., Gonzalez-Calero (2002). CBRonto: a task/method ontology for CBR, in: *Proceedings of the 5th International Florida Artificial Intelligence Research Society Conference*, May 14–16, Florida, USA.
- [11] V.T., D'Urso (2002). Home buyer search duration and the Internet. Working Paper 168, E-Business Center, MIT.
- [12] D. Ellis, A behavioral model for information retrieval system design, *Journal of Information Science* 15 (4/5) (1989) 237–247.
- [13] D. Ellis, D. Cox, K. Hall, A comparison of the information seeking patterns of researchers in the physical and social sciences, *Journal of Documentation* 49 (4) (1993) 356–369.
- [14] D. Ellis, M. Haugan, Modeling the information seeking patterns of engineers and research scientists in an industrial environment, *Journal of Documentation* 53 (4) (1997) 384–403.
- [15] D. Fensel, F. van Harmelen, I. Horrocks, D.L. McGuinness, P.F. Patel-Schneider, OIL: an ontology infrastructure for the semantic web, *Intelligent Systems* 16 (2) (2001) 38–45.
- [16] M., Fernandez-Lopez, A., Gomez-Perez, N., Juristo (1997). METH-ONTOLOGY: From ontological art towards ontological engineering, in: *Spring Symposium on Ontological Engineering of AAAI*, March 24–26, Stanford University, USA.
- [17] U. Flemming, S.-F. Chien, Schematic layout design in SEED environment, *Journal of Architectural Engineering* 1 (4) (1995) 162–169.
- [18] J.S. Ford, R.C. Rutherford, A. Yavas, The effects of the Internet on marketing residential real estate, *Journal of Housing Economics* 14 (2005) 92–108.
- [19] P. Gervas, B. Diaz-Agudo, F. Peinado, R. Hervás, Story plot generation based on CBR, *Knowledge-based Systems* 18 (4/5) (2005) 235–242.
- [20] I. Hsieh-Yee, Research on web search behavior, *Library & Information Science Research* 23 (1) (2001) 168–185.
- [21] B.J. Jansen, A. Spink, How are we searching the World Wide Web? A comparison of nine search engine transaction logs, *Information Processing and Management* 42 (1) (2006) 248–263.
- [22] Y.I. Kim, A study on the real estate retrieval system using web service, *Journal of Korean Institute of Information Strategy* 8 (2) (2005) 63–72. in Korean.
- [23] J.L. Kolodner, *Case-based Reasoning*, Morgan Kaufmann Publishers, 1993.
- [24] N. Kumar, K.R. Lang, Do search terms matter for online consumers? The interplay between search engine query specification and topical organization, *Decision Support Systems* 44 (1) (2007) 159–174.
- [25] A. Lau, E. Tsui, W.B. Lee, An ontology-based similarity measurement for problem-based case reasoning, *Expert Systems with Applications* 36 (3) (2009) 6574–6579.
- [26] H., Lebo (2003). Surveying the digital future – Year three. The UCLA Internet Report. Available at: <<http://www.digitalcenter.org/pdf/InternetReportYearThree.pdf>> (accessed by July 7, 2012).
- [27] V.Z. Leonard, H.J. Ken, I.A. Randy, Internet use and real estate brokerage market intermediation, *Journal of Housing Economics* 12 (2003) 134–150.
- [28] S.S. Liaw, H.J. Huang, An investigation of user attitudes toward search engines as an information retrieval tool, *Computers in Human Behavior* 19 (2003) 751–765.
- [29] F., Lorenzi, F., Ricci (2005). Case-based recommender systems: a unifying view, in: *Proceedings of the 2003 International Conference on Intelligent Techniques for Web Personalization (ITWP)*, August 11, Acapulco, Mexico.
- [30] Q., Yang, E., Kim, K., Racine (1997). Caseadvisor: Supporting interactive problem solving and casebase maintenance for help desk applications in: *Proceedings of the IJCAI'97 Workshop on Practical Use of CBR*, August 25, Nogoya, Japan.
- [31] J.A. Recio-Garcia, B. Diaz-Agudo, P. Gonzalez-Calero, A. Sanchez-Ruiz-Granados, Ontology based CBR with jCOLIBRI, in: R. Ellis, T. Allen, A. Tuson (Eds.), *Applications and Innovations in Intelligent Systems XIV*, Springer-Verlag, London, 2007, pp. 149–162.
- [32] P. Resnick, H.R. Varian, Recommender systems, *Communications of the ACM* 40 (3) (1997) 56–58.
- [33] J.B., Schafer, J., Konstan, J., Riedi (1999). Recommender systems in e-commerce, in: *Proceedings of the 1st ACM Conference on Electronic Commerce*, November 3–5, Denver, USA.
- [34] The 2011 National Association of REALTORS® Profile of Home Buyers and Sellers, Available at: <<http://www.carolinarealtors.com/files/National%20HBS%202011.pdf>> (accessed by July 5, 2012).
- [35] The 2009 National Association of REALTORS® Profile of Home Buyers and Sellers, Available at: <[http://www.abor.com/news\\_media/PublicAwareness/NARProfile.pdf](http://www.abor.com/news_media/PublicAwareness/NARProfile.pdf)> (accessed by July 10, 2012).
- [36] K.E. Weick, R.L. Daft, The Effectiveness of Interpretation Systems, in: K.S. Cameron, D.A. Whetten (Eds.), *Organizational Effectiveness: A Comparison of Multiple Models*, Academic Press, New York, 1983, pp. 71–93.
- [37] C., Williamson B., Shneiderman (1992). The dynamic HomeFinder: evaluating dynamic queries in a real-estate information exploration system, in: *Proceedings of the 15th ACM SIGIR Conference*, June 21–24, Copenhagen, Denmark.
- [38] S. Zheng, H. Liu, R. Lee, Buyer search and the role of broker in an emerging housing market: a case study of Guangzhou, *Tsinghua Science and Technology* 11 (6) (2006) 675–685.