Home suggestion service in real estate searching system

Using ontology and case-based reasoning

Yoonhyun Kim[1], Sangjin Park[2], Jieun Park[3], Sunjoong Kim[4], Jihyun Lee[5]
[1,3,4,5]Graduate school of Culture Technology, KAIST, Korea, [2]Graduate school of Knowledge Service Engineering, KAIST, Korea
[1]kimyoonhyun@kaist.ac.kr, [2]psj9920@kaist.ac.kr, [3]jieun315.park@gmail.com, [4] iuvenalis@hotmail.com, [5]jihyunlee@kaist.ac.kr

Abstract. *Real estate search system currently supports users' search for sale and consumption. However, the quality of the search results is unsatisfactory in most cases. This is due to the lack of system capacity to accurately understand and process all-different user preferences. Therefore, this study aims to solve the said issue by newly constructing real estate information via "Case-based Reasoning" and "Ontology." We conducted research on the terminology related to real estate search that could be seen in a number of Q&A communities. Based on the work, a hierarchical structure of the concepts was reorganized. In addition, a real estate search system was developed to authenticate its effectiveness while dealing with several exemplary cases.*
Keywords. *CaseBaseReasoning; Ontology; Information modeling; Service design; Knowledge acquisition.*

INTRODUCTION

Traditional way of home suggestion has relied on human realtor; however, the advent of the Internet has gradually been changing the long-established convention. Leonard (2003) insists that when it comes to retrieval of architectural information of real estate the Internet would play a significant role, increasing efficiency in and convenience of search. According to his study, it turns out that the use of the Internet in search of real properties decreases the total cost to find appropriate one as well as time, which emphasizes the necessity of digitized real estate information and online searching service.

Despite the expected advantages, however, digitization of real estate information is still in its infant stage and there are several problems in current practice. First, current web-based searching service needs to be more active and intelligent beyond online catalog in order to handle realty's special features such as immovability, localization, secrecy of deal and costliness and fluctuation of the price. Second, the information provided online is limited to basic features of real properties, which is not enough to lead customers to purchase decision(Kim, J.H, 2001 ; Lee, K.C. and Choi, K.Y., 2004). In addition, insufficient reliance on information and quality of suggestion cause user satisfaction to stay low in spite of the increasing use of online search(Cho, J.H. ,2009). Those problems are fundamentally caused by the absence of a systematic organization in which information is classified and connected to explain certain realty.

To supplement the low quality of home suggestion, this paper suggests integration of case-based reasoning with ontology. Case-based reasoning is a very useful paradigm to solve a current problem by referencing to the solutions to previous, similar problem. In other words, proposed searching system would suggest items chosen by previous users whose housing needs were similar with the current user. Consequentially, CBR would provide users with composed of more suitable real estate items to their needs and preference, which lightens user's effort, time, and cost to find a right house item. Furthermore, integration of CBR with ontology is expected to generate an incredible level of synergy by improving both the efficiency in information organization and effectivity of searched results.

The purpose of this research is to discuss the problems incurred by current practice for digitization of real estate information and to solve the problems by utilizing systematic organization and establishing CBR model in order for users to be offered appropriate result.

RELATED WORK

Current state of online real estate search service

Real estate search systems in Korea are mostly provided on the web. Analysis of real estate imformation providing service has already done in The development of a real estate multi-attribute integrated search system(Cho, J.H. ,2009). We refered to this paper to understand exsiting system to develope real estate search system. Web-based real estate search systems in Korea focus on 2 user groups, one is user of habitation purpose, anohter is user of investment purpose. For this reason, system has to be more complex, which cause difficulties to provide real estate information for living. Every real estate search system provide local-based search fucntion. System of "Budongsan 114" provide additional search fucntion such as station-based search, price-based search, and size-based search, but other limit to local based search. Although keyword based search fucntion is basic of systems and it is possible to choose additional 3 search options, still systems are lack of real estate recommend fucntion. Therefore, the development of a real estate multi-attribute integrated search system pointed out 4 problems;

1. Users feel difficulty to compare vest amount of data
2. User have to spend much cost, time and effort to find what he or she wants.
3. search fucntions are sequentially performed.
4. Limitation to access information for their vest amount of data.

Study combining case-based reasoning and ontology

Case Base Reasoning(CBR) is methodology to solve current problems based on previous case.(Kolodner, J.L. , 1992) Past expert system which solve the problem using IF – THEN rule based inference may need vest amount of rule, so it needs much time and cost to build Knowledge base. However, if we use CBR, we can find very similar solution using case, although there is no exact rules. Therefore, the more previous cases are stored, the better we can find more correct solution and relieve time and cost to build knowledge base. CBR system mainly used in E-commerce domain, such as Analog Device[1], WEBSELL(P. Cunnighsm, 2001), READEE.(P. Oehler, 1998). CBR system can infer more intelligently fitting for user's needs but it is hard to clearly specifiy the case. In the past CBR system, case is very restricted to data base table, so case can be restictly applied. In other words, we cannot descript relationship among various variables organizing the case, it is hard to provide flexible and semantically correct result.(Adela Laua et al, 2009). Recently, to give semanctic association among variables to solve current CBR problem, many researches on using ontology for CBR has rised. Ontology is suggested to enable quality automation of work including information search, intepretation, and integration. It avails humans and compturs of easily understanding each information by providing meanings to relationship among diverse, distributed information throughout the web.(T. Berners-Lee et al, 2001) Therefore, ontology has its advantage in systematically organizing caes, thus building up a knowledge model.

Table 1
Table title (Domestic real estate website search and comparison of analysis functions)

Website	Search Method	Location Analysis	Individual Analysis	Search Function
Budongsan 114	Provide information based on Location, Station, Size, and Price 2.Location is hierarchically provided 3.Station is provided in accordance with subway line 4. Size Price are provided with Location information	1.Provide Sub web site for each location 2.Autonomously provide market information for each location 3. Provide price information for each state, city, gu, dong	Size(exclusive area) Floor(relevant floor/total floor) the number of households the number of room (the number of bathroom) entrance structure(floor type/hole way type) price(sale price) tax information, construction company, construction year, heating system, parking availability, transportation educational facilities convenience feature photo information	Search function only available in recommended real estate and real estate of each dong price(range type)search size(range) search Arrangement function
	Feature: abundant data for real estate search and location-based analysis			
Speed bank	Location-based: gu / dong	- Provide Sub web site for each location -Apartmnet classification (gu/dong) Provide infrastructure and government building	Size(exclusive area) Floor(relevant floor/total floor) the number of households the number of room (the number of bathroom) Price(parcel price, premium, sale price), construction company, construction year, heating system, parking availability, transportation(subway, bus, road), educational facilities, convenience environment, feature, photo information	Select by types of property Select by types of deal(dealing, leasing, monthly rent) Search price(From highest price to lowest price) Search size(From least size to biggest size) Function of arrangement (Indescennding power and inascending power name of apartment, Size, number of floors, sale price, registration day, name of agent company) -Function of comparison(Maximum 5)
	Feature: Provide rich information in way of each analysis(price, traffic, green environment, photo information), superior search function			
Doctor Apart	Subdivision by each location Si(Do)-Gu-Dong	Provide autonomously Market information and Market price by each location Provide price by each Dong, price by each size, changing progress information in market price information	Size (exclusive area) Floor(relevant floor/total floor) the number of households the number of room (the number of bathroom) direction entrance structure(floor type/hole way type) price(sale price) construction company, construction year, heating system, parking availability, transportation(subway), educational facilities, convenience environment, feature,	Function of arrangement(name of apartment, size, number of floor, sale price, indescennding power and inascending power by registration day) -Function of comparision(Maximum 2)
	Feature: present market price information of each location with graphs			

METHODOLOGY

Definition and classification for construction informaton of real estate

Through the related work study, we investigated the definition and range of the construction information as well as the features that were used as information. Also, the analysis system was studied. We defined and classified apartment information by referring to the previous work including those of Kim Han Su(2001), Mun Tae Heon et al.(2008), Jung Sun O("1992), Heo Jin Seon et al.(2003), Kim Ta Yeol et al.(2000), Oh Chan Ok et al.(1994), Jung Hyeong Chul(2009), Lee Sae Young et al.(2006), Park Kang Cheol(1998), Kim Won Pal(1999), Ku Bon Chang et al.(2001), Choi Yeol et al.(2002), Lee Cheon Ki et al(2002), and Cho Sung Hee et al(1999).

Case-based construction method

In our study, we utililized top-down access to express real estate as cases.

Our study adopted the top-down approach and classified construction information. Then, sub-cases for each of the real estate cases were analyzed. Each case consists of different information due to their different features, so they must be partially separated. Therefore, we classified sub-cases differently in accordance with the characteristic of each real estate

1st Level	2nd Level	3rd Level	Information Provision
Unit Plan	Price	Parcel price, Current price	10,000 won(average price for maximum/minimum), Average sale price/pyeong
	Availability		Sale, price(monthly / yearly)
	Size	Size	Parcel price & size (pyeong,m^2)
		Room	Unit
		Bathroom	Unit
	Location	Residence	Floor
	Direction	Direction)	South, South-south-east, South-east, South-west, South-south-west, East, East-east-south, South-east
		Light	
		Vantilation	
		Noise	Closeness to the streets(over 30meters), not parallel to the streets, intensity of light, closeness to shopping districts
	View	Landscape, Openness	
	Structure	Multi-level structure, size of the room, hallway structure, exit method(stair vs hallway), valcony size & number	Parcel/room size
	Finishing materials and maintenance	Maintenance state, finishing material type, kitchen & bathroom facility	
	Interior renovation	Porch expansion & structure	
	Welfare facility accessibility	Superintendent's office, playground, amenities, fitness clubs, accessibility to parking space	

Table 2
Table title (Information classification of apartment information)

1st Level	2nd Level	3rd Level	Information Provision
Block information	Development Feature	Main agent(construction company)	Private/official, designated/undesignated
		Development method	Novel development, renovation, redevelopment
	Duration	Completion duration	Year
	Location	Address	Autonomous region, dong
		Distinction	Downtown, North-west, North-east, South-west, South-east
	Block Size	Total family, floor, size, distance	
	Housing structure	Hallway-based, stair-based, multi-purpose	
	Economic feasibility	Average maintenance fee	
	Heating	Heating method	Individual, central
		Heating source	LPG, oil. gas
	Service elements	Move-in & parcel-out service	
		Guard	
		Crime prevention	
	Block life	Garbage disposal	
		Parking	The number of cars/family
		Welfare & amenities	Playground, senior citizens' center
		Green area	Landscape size
		Access way	Block main exit
Location Information	Convenience Environment	Distance to public or government office, hospitals, markets, etc / The number of bus channels or operations, etc	
	Education Facility Environment	Distance to elementary, middle, or high schools or colleges/ The number of middle-schools	
	Park space	Distance to green space	
	Transportation	Distance to downtown or subway	
	Inconvenience	Distance to industrial complex, noise source, or stench source	

case, and tried to compose necessary, relevant information within each case.

Ontology construction under methodology formats

Methontology , one of the several methods for ontology construction, proceeds with ontology knowledge model construction at the developmental phase of the software engineering. (Fernández-López M, 1997 ; Oscar Corcho1, 2005) Methondology can function as a significant basis for ontology knowledge model construction since it takes general elements into consideration rather than leaning toward

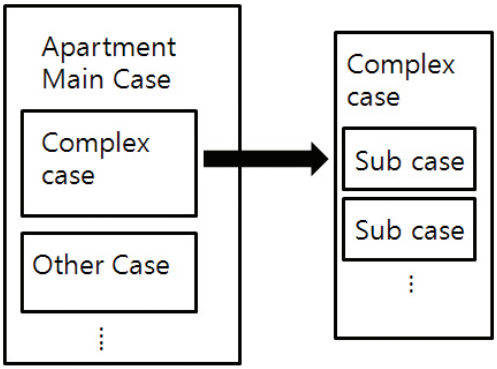


Figure 1
{Top-down access to concepts for Korean habitation cases}

a specific domain. Therefore, this study aims to utilize methontology model while individually considering our domain specific features at the same time.

Conceptualization

We developed an abstract framework for the ontology construction while organizing keywords and analyzing user requirements in the websites. Using the framework, we plan to compose concept dictionary, concept classification tree, relationship table, and feature table. Meanwhile, we will continuously evaluate the usability of diverse virtual and real scenarios.

Hierarchical structural organization for keywords

First, we investigated into domestic and international Q&A websites to collect user Q&A data used for actual real estate searches. The representative web sites of Yahoo! Answers, WikiAnswers, and Naver KnowledgeiN was chosen for the analysis, and keywords of "real estate" and "recommend" was used. We analyzed necessary terminology for ontology construction and organized the concepts utilizing knowledge acquirement and scenario analysis results.

Classification of Concepts and Definition of Relationships

A work for constructing ontology knowledge framework for future inference was conducted while organizing and classifying keywords and concepts obtained from Table 3.

Combination of case and ontology

The advantanges of the ontology on this study can be classified as follows.

First, if elements including location and place information that should be essentially included in the real estate information are integrated into ontology, management efficiency of information will be heightend while solving data-overlap-related issues.

As a basis for the first propsal, unncessary data overlapping occurs when cases are constructed under top-down method. For instance, if apartment information is built up under apartment case formats with sub-set cases, than overlap will occur. Therefore, we utilized ontology in this study in order to solve

Table 3
{Concept classification through keyword analysis}

Upper concepts	Included concepts
Location	Seoul, Daejeon, Daegu, Guri, Kayadong, Daeyeondong, Yongsangu, Kangnam, etc.
RentType	Deposit, Lease (yearly/monthly), Sale, etc.
ArchitectureType	Apartment, Single-room, Villa, Single-house, Lease apartment, Office-tel, etc.
Environment	Station, University, Highschool, Mart, Cultural infra, Job-place, Green area, Walking path, Park, Fresh environment, etc.
Price	30 million won, 1 million won, 1.5 million won, 100 thousand won, 60 million won, etc.
Size	35Pyeong, 25 Pyeong, 30 Pyeong, etc.
Room	Two rooms, Three rooms, One bathroom, Two bathrooms, Living room, etc.
Inner Structure	Interior, draft, etc.
Estate	Whole real estate

the said issue. There are common, essential elements for each real estate such as location and place information. Such elements should be individually constructed and synthetically managed for data overlap prevention. So to speak, if location ontology can be individually constructed for location-related information, than information overlapping problems regarding location can be reduced.

As a basis for the second proppsal, cases are composed of sub-set cases that comprise of each case. Therefore, if relationships among sub-cases are defined and related information managed by indexing function of ontology, than overall management of the information will become much easier and more efficient.

IMPLEMENTATION

The following is the model that combines case and ontology. OWL is separated into OWL Lite, OWL DL, OWL Full according to the degree of expressiveness. (Donghee Yoo and Yongmoo Suh, 2008) Since current inference engines only guarantees decidability of OWL Lite and OWL DL, OWL DS was utilized that owns better expressiveness out of the two. In order to materialize ontology Protégé v3.4.4 [2] developmental tool was used. Figure 6 was expressed as ontology while using apartment cases and sub-set cases.

Common information of the real estate such as location information was shared within ontology structures in order to prevent data overlapping. (see Figure 7)

The following table indicates organized cases of apartments. Class and property information for the apartment cases is organized.

System Architecture

As seen in Figure 7, the system architecture of the real

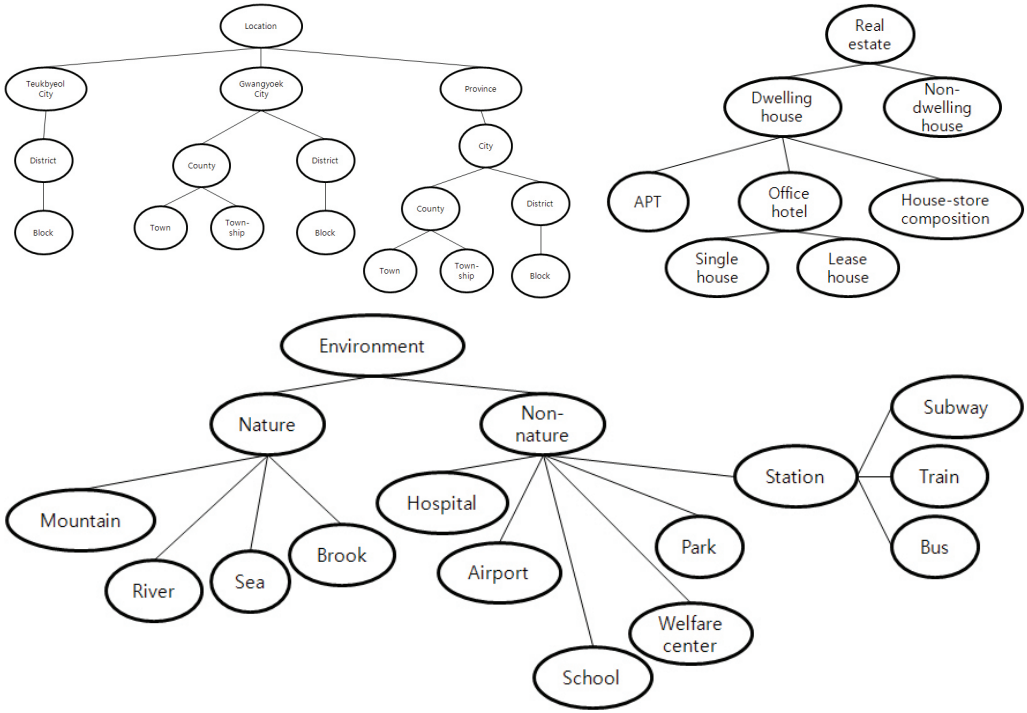
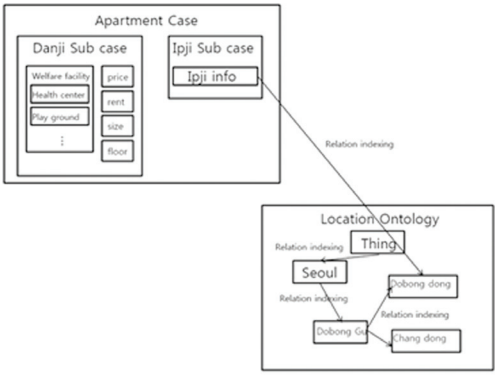


Figure 2
{Hierarchical structure for concepts of location, real estate, and environment}



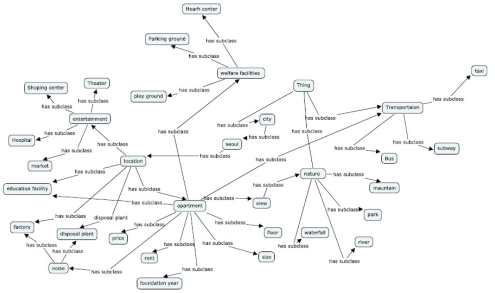
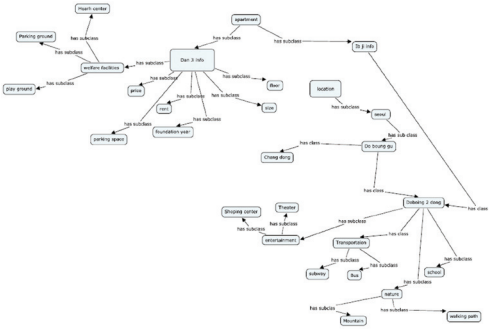


Figure 3 (left)
{Prevention of data overlapping by using ontology}

Figure 4 (right)
{Conceptual diagram for information relationship}



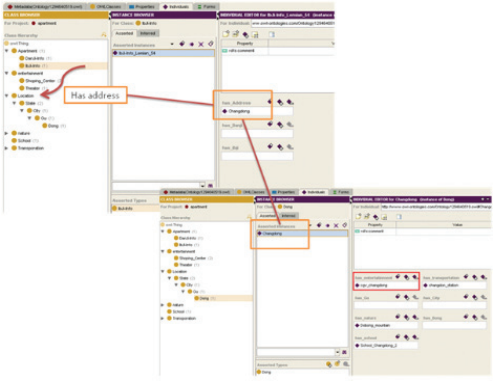


Figure 5 (left)
{Integration of case and local-ontology}

Figure 6 (right)
{Ontology construction utilizing Protege3.4.4}

Ontology	Constitution	Concept
Apartment	Class	Ipji-info, Danji-info
	Object Property	Has_danji, Has_ipji
Danji-info	DataType Property	Has_floor, has_foundation_year, Is_rent, health_center, Has_parking_space, Has_play_ground, Has_price, Has_size
Ipji-info	Object Property	Has_address
Entertainment	Class	Shoping_Center, Theater
Nature	Class	Mountain, Park, Walking_path
School	DataType Property	Has_highschool, Has_middleschool, Has_elementryschool
Transportation	Class	Bus, Subway
Location	Class	State,City,Gu,Dong
	Object Property	Has_city, Has_gu, Has_dong
Dong	Object Property	Has_school, Has_transportation, Has_nature

Table 4
{Ontology concept related to apartment real estate}

Figure 7
{System Architecture}

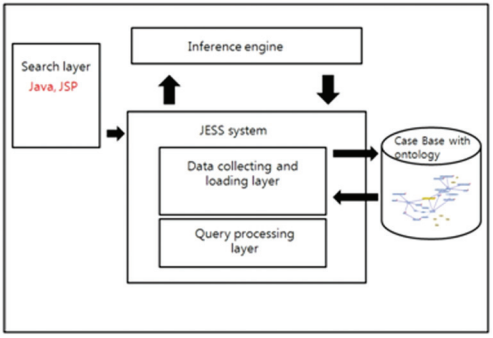
estate search systyem consists of 1) Search layer, 2) Data collecting and loading layer, 3) Inference layer, 4) Casebase with ontology, and 5) Query processing layer. The main functions of each layer is as follows.

- Search layer: The main screen for real estate related search of users
- Data collecting and loading layer: Delivers real estate information from case-based to inference. Collected and inferred information is then again saved in the case-based.
- Inference layer: Executes inference on real estate information and delivers inferred results to data collection and loading layer when new information is being inferred.
- Query processing laye: Analyzes user questions obtained from the search layer, and supports data loading layer to acquire necessary information.
- Casebase with ontology: Database where case-based real estate information and ontology are combined.

CONCLUSION

The common problematic issues that current real estate search system confront include insufficient search options, difficulty of comparative analysis throughout vast information, large expense of money for quality search, great deal of time consumption, limitation of linear order of search, and constraint access to information. In order to solve the issues, this study proposed to combine detail subsets of case information into ontology while builiding up cases for the real estate information. Its implications in relation to the real estate search can be summarized as follows.

First, a platform for execution of diverse methods of real estate search was constructed since a great amount of objective terminology was defined. Also, the concepts that individuals differently perceive was organized as cases, resulting in quality search of the users who required personalized, relevant data. Last, it successfully developed a real estate system that enabled enhanced quality search based on ontology. As future work, a service



that allows more personalized real estate search should be developed while utilizing individual, cognitive ontology for conceptual construction. Furthermore, usability testing must be conducted to evaluate how users perceive ease of system use or learnability of our prototype.

REFERENCES

Laua A, Tsuib E and Leeb WB 2009 'An ontology-based similarity measurement for problem-based case reasoning', *Expert Systems with Applications*, vol.36, no.3, pp.6574-6579.

Cho, JH 2009 'The development of a real estate multi-attribute integrated search system', *The Journal of Korean Institute of CALS/ES*, 14(3), vol.14, no.3, pp.15-37.

Donghee Y and Yongmoo S 2008 'An Ontology-based Hotel Search System Using Semantic Web Technologies', *Society for e-business studies*, 14(3), vol.13, no.4, pp.71-92 .

Fernández-López M, Gómez-Pérez A and Juristo N 1997 'METHONTOLOGY: From Ontological Art Towards Ontological Engineering', *Spring Symposium on OntologicalEngineering of AAAI,* Stanford University, California, pp. 33–40.

Kim, JH 2001 'A Study on Agent System for Personalized Real Estate Information Services', *M.S. thesis,* Han-Yang University, Seoul, Republic of Korea.

Kolodner, JL 1992 'An introduction to Case-Based Reasoning', *Artificial Intelligence Review*, vol.6, no.1, pp.3-34.

Lee, KC and Choi, KY 2004 'Development of evaluation model for real estate web sites' *Korea Real Estate Academy*, vol.23, pp.35-52.

Leonard, VZ, Ken, HJ and Randy, IA 2003 'Internet use and real estate brokerage market intermediation', *Journal of Housing Economics,* 12(3),vol.12, no.2, pp.134-150.

Corcho1 O, Fernández-López M, Gómez-Pérez A and LópezCima A 2005 'Building legal ontologies with METHONTOLOGY and WebODE', *Law and the Semantic Web Lecture Notes in Computer Science*,vol.3369, pp.142-157.

Cunnighsm P 2001 'WEBSELL: Intelligent sales assistant for the world wide web', *Kunstliche Intelligenz (KI)*, Arendtap, Bremen, Germany.

Oehler P, Vollrath I, Conradi P, Bergmann R and Wahlmann T 1998 'READEE – Decision Support for IP Selection using a Knowledge-Based Approach', IP 98. *Europe*, Miller Freeman, San Francisco, CA.

Berners-Lee T, Hendler J and Lassila O 2001 'The Semantic Web', *Scientific American*, pp. 28-37.

[1] http://analog.com

[2] http://protege.stanford.edu