MyINS: A CBR e-Commerce Application for Insurance Policies

SITI SORAYA ABDUL RAHMAN, AZAH ANIR NORMAN and KOH JOE SOON
Faculty of Computer Science and Information Technology
University of Malaya
50603 Kuala Lumpur
MALAYSIA
siti_soraya@um.edu.my, azahnorman@um.edu.my, joe_spoon@yahoo.com

Abstract: - This paper presents the design and development of an insurance policies recommender for prospect client. The main aims of MyINS are twofold. The first aim is to simplify the process for client in making decision base on the information recommended through this system. The second aim is to allow client to choose the insurance policy that is most suitable for them. MyINS models the reasoning process employed by insurance sale agent in proposing policy to the client using case-based reasoning algorithm (CBR). CBR, generally describe the process of solving current problems based on the proposed solution from similar past problems. MyINS works with a principle assumption that different client with different background will have different type of insurance plan suitable to each one of their needs. MyINS makes recommendation of insurance policies based on personal data and desired coverage of the policyholder from the similar past problems. Similarity assessment and demonstration of case library are also presented.

Key-Words: - Insurance Policy application, online application, e-commerce solution, Case-Based Reasoning algorithm, Web based

1 Introduction

The Internet has offer great business opportunity for entrepreneurs to expand their business internationally. According to Elias Awad (2004) [1], e-commerce brings the universal access of the Internet to the core business processes of buying and selling goods and services. It enable business to achieve goals such as reaching to new markets, creating new products and services, building customer loyalty, enriching human capital with vigorous application and skill developed by the new system, making use of the existing and emerging technologies and last but never the least is creating a competitive advantage which is essential for all business in the world. New e-commerce model (in broad categories) shows different types of taxonomy such as business-to-consumer, business-to-business and business-to-government as discussed by Timmers (1998) [2] and Kalakota and Whinston (1996) [3]. These types of models are adopted by the businesses to suit their objectives and needs. As discussed by Peter Lovelock and John Ure, the director and deputy director respectively of the Telecommunications Research Project, University of Hong Kong (www.trp.hku.hk) [4], explains that successes of the e-commerce are mainly due to time, disappearance of geographic and economic boundaries, disintermediation, open source and the catalytic effect e-commerce comprise of, whereby it has change the organisation work structures. The e-commerce successes also are due to the increase of interactivity and the ability to create numbers of channels for knowledge diffusion in the workplace.

JingTao Yao, in one of her articles in the Journal of Electronic Commerce Research, VOL. 5, NO. 1, 2004 quoted from Grace (1998) [5], that there is no exception for insurance business that is currently experiencing a transformation with technology, an industry where electronic commerce will play a significant role. The development of search engines and in particular the intelligent agents or software agents have attracted more businesses including the insurance company to participate in e-commerce. Efraim Turban (2006) [6] explains that these software agents are used to support tasks such as comparing prices, interpreting information, monitoring activities and working as an assistant, where user can chat and collaborate with the agents. This resulted in proliferation of hundreds of companies to offer online sales and services including the insurance companies' which offer websites that have online insurance sales and services which are distributed via the Internet. E-commerce brings automation and digital-online delivery toward the insurance business' application and the system. Nowadays, clients can directly obtain insurance quotes and purchase policies faster and easier from insurance company website, as the clients now play more active role in selecting and analyzing the right policy that is best for them. The dedicated websites are designed to assist prospect client in making an intelligent decision. The only problem is that, such online quotes are rarely understood by a prospect client. In addition, the insurance policies quoted cover small amount of client's important aspects, example health and lifestyle. In this paper, the discussion will present an approach for selecting insurance policies that makes use of case-based reasoning algorithm. Following this approach, each policyholder is represented as a case. The new case is solved based on the proposed solution of similar past problems (old case) the system has encountered.

According to Luger and Stubblefield (2005) [7], CBR is another powerful strategy experts use to address new problem solving solution. Zhaohao, Finnie and Weber (2004) [8], quoted that the goal of CBR is to infer a solution for a current problem description or enquiry in a special domain from solutions of a family of previously solved problems, the case library. CBR systems are a particular type of analogical system which has an increasing number of applications in different fields such as in intelligent Web-based sales services and Web-base planning as well as multi-agent systems (Zhaohao, Finnie and Weber, 2004) [8]. In general, there are 4 RE processes involved in a CBR cycle (Nakasumi, 2003) [9] which include:

1. RETRIEVE the most similar case or cases
2. REUSE the information and knowledge in that case to solve the problem
3. REVISE the proposed solution
4. RETAIN the parts of the experience likely to be useful for future problem solving.

In current traditional practice, CBR systems are loaded in the PC and run under the individual operating system. This creates inconvenience to the users and significantly influences the company in terms of creating, disseminating, storing, and retrieving paper-based information and this leads to high operation cost. The benefits of the traditional CBR can be further leveraged by porting the CBR on the Internet. Web-based CBR offers several advantages over the traditional CBR in four ways: (1) Accessibility: where the Internet permits unlimited access to web-based CBR system. (2) Interoperability: in which the Internet permits a single shared web-based CBR system across different platforms. (3) Transferability: whereby web-based CBR system offers new way of delivering services and distributing expert advice over the Internet which acts as sales assistant for E-Commerce. (4). Maintainability: the web-based CBR system cuts cost of the system maintenance since there is only a single, centralized case library. According to Watson (2000) [10] as quoted from Hammond et al. (1996), work on web-based CBR systems have been established for a few years such as the FAQFinder and FindME systems.

2 MyINS

The following subsection further explains MyINS architecture and CBR methodology.

2.1 MyINS Architecture

MyINS is a web-based application. The users can access to the system at anytime and any place, wherever an Internet connection is available. The system is based upon client/server three-tier architecture. The three layers include client tier, application tier and data tier. The client web browser supports MyINS system interface. The server software in the application tier supports the reasoning mechanism of MyINS system. This server-side reasoning capability is programmed using C# scripting language and developed over Microsoft Windows XP Professional platform. The server-side case library in data tier plays a vital role in the system development. It is used for storing cases which are needed for insurance policy recommendations. The case library is designed using MySQL Server Database. Figure 1 shows the three-tier web application architecture for MyINS system.

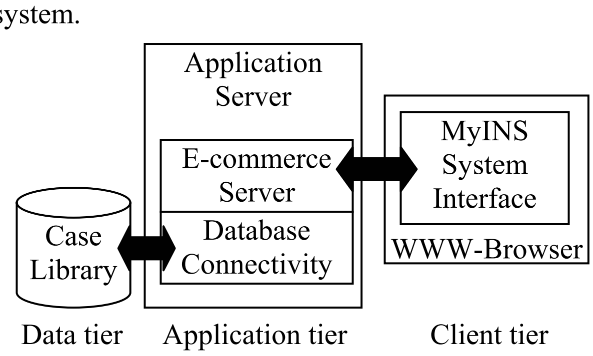

Fig. 1: Three-tier MyINS Architecture

Based on the three-tier MyINS architecture above, communication between the client tier to the data tier are detailed in the descriptions below:

a) User interacts with the web browser (Internet Explorer, Mozilla, or Netscape Navigator) which contains MyINS system interface in the form of a HTML page.
b) C# scripting is embedded in the HTML page using ASP.NET.
c) User inputs all information required by MyINS system in the forms presented in MyINS system interface seen in the client browser. This is a data collection process for MyINS system. This process resides in the client tier.
d) The user's web browser sends requests for data to application tier.
e) The application tier receives data from the client tier.
f) The Microsoft Internet Information Services web server recognizes the requested file in ASP scripting format, and interprets the file, before responding to the page request.
g) The request is evaluated and passed to the data tier.
h) Certain SQL query commands will connect to the Microsoft SQL Server database and request the content that belongs in the web page.
i) Microsoft SQL Server database responds by sending the requested content to ASP.
j) The C# script stores the content into one or more C# variables, and then output the content as part of the web page.
k) The ASP finishes up by handing a copy of the HTML it has created to the web server.
l) The web server sends the HTML page to the web browser as a plain HTML file. The page output is provided by the ASP scripting format.

2.2 MyINS CBR methodology

The system developed in this project has been divided into four major modules. Each of the modules performs different tasks. Figure 2 illustrates MyINS CBR model.

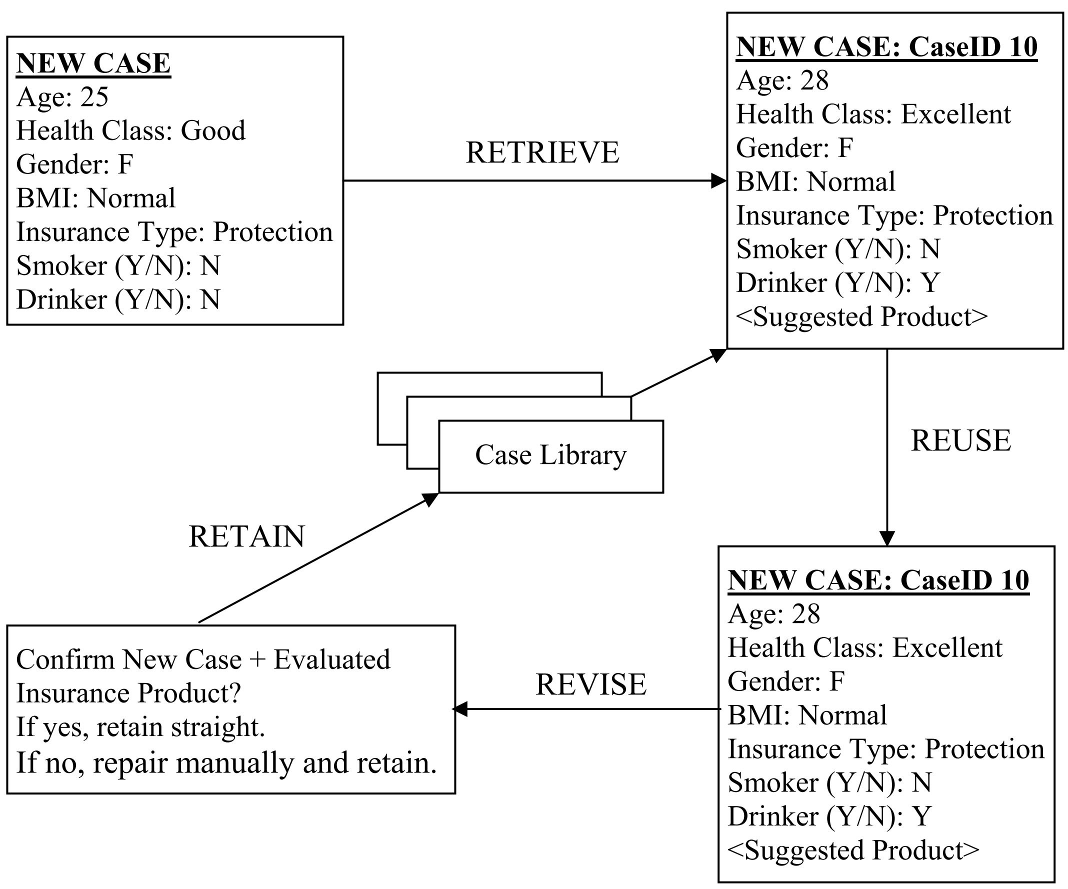

Fig. 2: MyINS CBR model

2.2.1 RETRIEVE module
In CBR, the similarity of cases in the case library to a query case is defined by similarity function. MyINS used Nearest Neighbour algorithm to computes the similarity of the query case to all the cases in the case library based on weight features. Section 4.1 further explained Nearest Neighbour algorithm employed in MyINS. Having computed the similarity, MyINS retrieves cases in the case library with similarity scores that exceed a given threshold for reused module.

2.2.2 REUSE module
The reuse phase in MyINS simply returns the old solutions of the retrieved cases unchanged, as the proposed solution for the query case. According to De Mántaras (2006) et al [11], reuse becomes complex when there are significant differences between the query case and the retrieved case. If the case falls under this scenario, the retrieved solution might need to be adapted to account for the differences. [11] further states, adaptation become particularly relevant when used for constructive problem-solving tasks such as design, configuration, and planning. In MyINS, it is likely that the exact solution of the retrieved case be applied directly to the target problem.

2.2.3 REVISE module
At this revise phase, the solution is evaluated. The revise phase provides a means for the system to learn from failure when a solution generated is incorrect or a close match. Thus, requires necessary repair to the case solution. MyINS repairing solution is done manually by the domain expert.

2.2.4 RETAIN module
Case is retained based on the final confirmation from the user. If the solution is an exact match and accepted by the user, it will be retained as a new case in the case library. If the solution is not accepted, user will be prompted to either discard the solution proposed or restart the process again.

3 Case Example
Kolodner (1993) as adapted from Binh V. Nguyen (2002) [12], defined a case as "a contextualized piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goal of the reasoner". [12] describes three major parts of a case include the problem descriptions, the solution and the outcome. Below are the three major parts of MyINS case.

3.0.1 The problems description
In MyINS, the problem description is the list of features describing policyholder in the case. MyINS system uses flat feature-value lists as its case representation technique. The system has specific features where each feature is with different weights assigned to each of the feature values. The weights allow specific features to have different importance in terms of its value. The weights are supplied by the knowledge engineer and the domain expert. Table 1 shows the samples of the features of a case along with its values, remarks and its weights. (Note: The weights assigned to each of the features are according to how qualified or how common the value compared to the others. The total weight value for each feature is 1.0)

3.0.2 The solution
The solution in MyINS is to introduce the right insurance policy.

3.0.3 The outcome
In MyINS, the outcome is the list of right insurance policy recommendations for prospect clients. The lists of right insurance policies are derived by retrieving past cases of policyholders which has similar problem and reusing them to current problem.

3.1 An example case
The following depict an example of a case in MyINS system. Each case consists of problem description with appropriate product recommendation useful for future problems.

Problem description
Age: 28
Gender: F
Smoker (Y/N): N
Health: Excellent
Policy: Protection
BMI: Normal
Drinker (Y/N): Y
Recommendation: Product 1

Features	Values	Remarks	Weights
Age	Classification: Juvenile: 0<age<=18 Young Adult: 18<age<=40 Middle Age: 40<age<=55 Old: 55<age<=70	General Insurance knowledge: Young – Low premium & high return Old – High premium & Low return	Juvenile = 0.4 Young Adult = 0.3 Middle Age = 0.2 Old = 0.1
Gender	Male or Female	Female has a little more privileges to certain products.	Female = 0.6 Male = 0.4
Smoker	Either the applicant is a smoker or non-smoker (within last year period)	Smoker pays higher premiums.	Non-smoker = 0.7 Smoker = 0.3
Health	Excellent, Good, Poor	A person with poor health or has critical illness may need to pay a higher premium or may not qualify for certain policies.	Excellent = 5 Good = 3 Poor = 0.2
Policy	You can choose either Protection policies or Health policies	Protection protects from unexpected events and hardships and also acts as a value builder too. Health reimburses medical expenses incurred due to accidents or illness.	Protection = 0.6 Health = 0.4
Body Mass Index (BMI)	Classification: Normal Underweight Overweight Obesity	It is a way of identifying the category of a person's body mass. BMI = weight/(height*height)	Normal = 0.4 Underweight = 0.3 Overweight = 0.2 Obesity = 0.1
Drinker	Either the applicant is a alcoholic or non-alcoholic drinker	Drinker pays higher premiums.	Non-drinker = 0.6 Drinker = 0.4

Table 1: MyINS case representation

4 MyINS Case Library

MyINS system uses the nearest neighbour algorithm for retrieval of insurance policy. The cases are stored using flat feature-value lists case representation technique.

4.1 The reasoning

As described in Table 1, each case stored in MyINS system is described as a set of features. The similarity between an input and a retrieved case stored in case library can be defined as the weighted sum of the similarity between corresponding cases features. The higher the similarity score, the more similar two cases are. A typical equation for similarity computation is shown below:

$$\frac{\sum_{i=1}^{n} W_i \times sim(f_i^i, f_i^R)}{\sum_{i=1}^{n} w_i} \qquad (1)$$

Where w is the weight for the feature, the higher the value of w, the more important the feature is. *sim* is the similarity function, and f^I and f^R are the values for feature i of the input case and retrieved case respectively.

4.1.1 Similarity for numerical features

To compute the similarity for numerical features, the approach is to start with a Distance Function. Distance Function is defined as:

$$d_{f(input,retrieved)} = |i - r| / range \quad (2)$$

$$Similarity_{f(input,retrieved)} = 1 - d \quad (3)$$

Based on the equation (2) and (3) above, the following example depicts the derivation of similarity metric for numerical feature (i.e. age) between the input value and the retrieved value.

Input case (New Problem): Age = 21
Retrieved case: Age = 42
Range (Age) = 0…70 years old

$$d_{Age(input,retrieved)} = |21 - 42| / 70 = 0.3$$

$$Similarity_{Age(input,retrieved)} = 1 - 0.3 = 0.7$$

The similarity measure is to score 1 for feature with equal value between input case and retrieved case.

4.2 CBR scenario example: The case of MyINS

Example: Input case is the new target problem and Case ID 12 and 33 are one of the stored cases in the case library. And the similarities between the features respectively have been calculated using the equation (2) and (3).

Input Case		Case ID 12
Age: 21	0.7	Age: 42
Gender: Female	1.0	Gender: Female
Smoker (Y/N): N	0.6	Smoker (Y/N): Y
Health: Excellent	0.8	Health: Good
Policy: Protection	1.0	Policy: Protection
BMI: Underweight	0.9	BMI: Normal
Drinker (Y/N): N	1.0	Drinker (Y/N): N
		Recommendation Product 1

$$Similarity\ (New, CaseID:12) = \frac{(0.3*0.7+0.6*1.0+0.7*0.6+0.5*0.8+0.6*1.0+0.3*0.9+0.6*1.0)}{(0.3+0.6+0.7+0.5+0.6+0.3+0.6)}$$

$$= 3.1/3.6 = 0.86$$

Note: Numbers in brackets are the feature's weight values found in the case library.

Case ID 33 shows higher similarity value (with a given similarity threshold) than Case ID 12, therefore Product 2 is recommended.

Input Case		Case ID 33
Age: 21	0.87	Age: 30
Gender: Female	1.0	Gender: Female
Smoker (Y/N): N	1.0	Smoker (Y/N): N
Health: Excellent	0.8	Health: Good
Policy: Protection	1.0	Policy: Protection
BMI: Underweight	0.9	BMI: Overweight
Drinker (Y/N): N	0.8	Drinker (Y/N): Y
		Recommendation Product 2

$$Similarity\ (New, CaseID:33) = \frac{(0.3*0.87+0.6*1.0+0.7*1.0+0.5*0.8+0.6*1.0+0.3*0.9+0.6*0.8)}{(0.3+0.6+0.7+0.5+0.6+0.3+0.6)}$$

$$= 3.31/3.6 = 0.919$$

5.0 System Demonstration

MyINS first step requires acquiring the user personal data (i.e. case query) from the case library. The second step involves the uses of nearest neighbour algorithm to help MyINS to retrieve the most similar case from the case library. Once data is computed, MyINS will then display the list of recommendations for the right insurance policy for each prospect user. If prospect user is happy and accept the recommendation, MyINS will then store the case (computed for the particular prospect user) to the case library. The system demonstrates that the interaction and transaction of data are according to the three tier architecture discussed previously. All cases computed are required to be stored in the case library as a carbon copy to ensure that the most similar cases could be retrieved for the usage of other prospect user. Hence it will offer solid references for the system to compute and recommend the best policies to future prospect users.

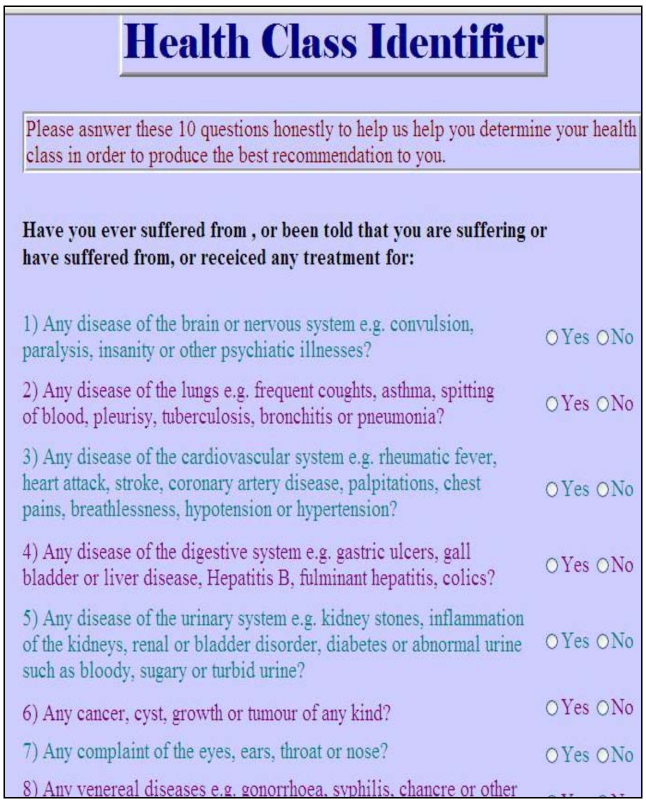

Fig. 3: Health Class Identifier page

Figure 3 shows the Health Class Identifier page where short questionnaires with 10 simple questions are asked to help find out the health class of the user. This Health Class Identifier page is a required page to track the users' health class in order to give better recommendation of insurance policy.

Fig. 4: Main Form page

Figure 4 shows Main form page of MyINS system. This page acquires users' personal data as a new target problem to solve.

CaseID	Health Class	Age	Body Mass Index	Gender	Policy	Smoker	Drinker	Similarity (x100%)
2	Good	30	Normal	Male	Protection	Yes	No	0.9608465
13	Good	28	Normal	Male	Protection	Yes	No	0.9576719
10	Good	43	Normal	Male	Health	Yes	Yes	0.9444444
4	Good	26	Obesity	Male	Health	Yes	No	0.9396825
27	Poor	45	Underweight	Female	Health	Yes	No	0.9301587
37	Excellent	23	Normal	Male	Protection	Yes	No	0.9275132

Fig. 5: Results from MyINS session

Figure 5 depicts the result from MyINS recommendation session which includes the percentage of similarity score above given threshold.

6 Conclusions and Future Work

MyINS is a proof-of-concept prototype of applying CBR algorithm in an insurance industry. The methodology presented in this paper may serve as a generic framework and can be used for other domains. However, there are many aspects of the prototype introduced in this paper deserve more attention in future research. Further research may include discussion on items such as:

1) Introducing more features for a case to improve the result of system in recommending the right insurance policy thus better reuse of existing case.
2) Combining two or more retrieval strategy to lead to high quality to retrieved solution.
3) Building case bases in the case library using a form of similarity based reasoning suggested by [8], with the objective to reduce the case library size to only most appropriate cases.

References:

[1] Elias M. Awad, *Electronic Commerce: From Vision to Fullfilment*, Prentice Hall, 2004

[2] Timmers, P, *Strategies and models for business-to-business trading*, New York: John Wiley ,1998.

[3] Kalakota, R. and Whinston, A.B., *Frontiers of Electronic Commerce*, Reading, Mass, Addison-Wesley, 1996.

[4] Peter Lovelock and John Ure, *The New Economy: Internet, Telecommunications and Electronic Commerce?*, September 2006 URL: http://www.trp.hku.hk/publications/new_economy.pdf

[5] JingTao Yao, *Ecommerce Adoption of Insurance Companies in New Zealand* Journal of Electronic Commerce Research, VOL. 5, NO. 1, 2004, pp 54-61

[6] Turban Efraim et al. *Ecommerce 2006: A Managerial Perspective* New Jersey Pearson Prentice Hall, 2006.

[7] Luger G., Stubblefield A. W. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Addison-Wesley, 2005.

[8] Zhaohao S., Finnie G., Weber K, *Case Base Building with Similarity Relations*, Information Sciences, An International Journal VOL. 165, 2004, pp 21-43 September 2006 URL: http://www.sciencedirect.com/science?_ob=MImg&_imagekey=B6V0C-49XPCKB-3-7D&_cdi=5643&_user=2980760&_orig=search&_coverDate=09%2F03%2F2004&_sk=998349998&view=c&_alid=453254243&_rdoc=1&wchp=dGLbVtz-zSkzk&md5=222e7e8dd6b4b43656aa86c4850f948b&ie=/sdarticle.pdf

[9] Nakasumi M, *Credit Risk Management System on e-Commerce: Case Base Reasoning approach*, ACM International Conference Proceeding Series, Proceedings of the 5th International Conference on Electronic Commerce; Vol. 50, 2003, pp 438-449, September 2006 URL: http://delivery.acm.org/10.1145/950000/948062/p438-nakasumi.pdf?key1=948062&key2=2750388511&coll=&dl=ACM&CFID=15151515&CFTOKEN=6184618

[10] Watson, I. *A Case-Based Reasoning Application for Engineering Sales Support using Introspective Reasoning.* In, Proc. 17th. National Conference on Artificial Intelligence (AAA!-2000) and the 12th Innovative Applications of Artificial Intelligence Conference (IAAI-2000), Austin Texas, AAAI Press. ISBN 0-262-51112-6, July 30 - August 3, 2000, pp. 1054-1059.

[11] De Mantaras et al. *Retrieval, reuse, revision and retention in case-based reasoning.* The Knowledge Engineering Review, Cambridge University Press, United Kingdom Vol. 20:3,.2006, pp 215–240

[12] Binh. V. Nguyen, *AAA: Another Academic Advisor using Case-Based Reasoning.* Ohio University. September 2006 URL: http://www.cs.umd.edu/~bnguyen/papers/aaa.pdf