

## A Web-based Game-Oriented College Selection System Employing Fuzzy Rule Trees

Sushil. K. Prasad, Raj Sunderraman, and Yanqing Zhang  
Department of Computer Science  
Georgia State University  
Atlanta, GA 30302  
[spasad@gsu.edu](mailto:spasad@gsu.edu)

Atul Parvatiyar  
Department of Marketing  
Georgia State University  
Atlanta, GA 30302

### Abstract

*Our web-based interactive game-playing-oriented college selection system acts as an smart advisor/mentor and helps students, parents, and teachers use an effective graphical user interface to efficiently search college information and to make good decisions at each stage of students' academic development. It is an expert agent with hierarchical fuzzy knowledge base using fuzzy logic. Users are expected to be able to use this agent as a trusted, involved and smart counselor, who remembers their profile and their past history and advises them through their high school years. It gives them increasingly better and specific nuggets of advice as they make progress through their schools, and guides them toward a selected set of colleges to apply that optimizes their potential. At the heart of this system is an assessment tree, which uses bottom-up fuzzy calculations to generate possibility of admission in various sets of colleges. The primary output is a short list of colleges to apply containing five kinds of colleges (highly-selective down to non-selective) with possibilities of acceptance in each college according to fuzzy rules provided. It is also a smart tool to play self-guided "what if" games to explore what specific action one should be taking to have maximum impact on the selection of colleges one can get in and on his/her chances of admission. It employs a natural interactive interview process for a user (i.e., a user can play a game with the agent to gradually find out rational solutions), and produces advisories at different stages.*

### 1. Introduction

To promote the use of fuzzy logic in the Internet applications, BISC International Workshop on Fuzzy Logic and the Internet was held at UC Berkeley, and a BISC Special Interest Group on Fuzzy Logic in the Internet was formed in 2001. Zadeh highlighted: "fuzzy logic may replace classical logic as what may be called the brainware of the Internet." Computational Web Intelligence (CWI), a new research area, was discussed at the special session at World Congress on Computational Intelligence (CI) in 2002. CWI is a hybrid technology of CI and Web Technology (WT) dedicated to increasing Quality of Intelligence (QoI) of e-applications on the Internet and wireless networks [3]. Fuzzy Web Intelligence (FWI), one technology of CWI, focuses on using fuzzy theory and Web technology to develop intelligent Web application systems. The smart college agent (SCA) expert system in this paper is a good example of FWI technology. The basic architecture of the college selection system was introduced in [8]. The design of hierarchical fuzzy expert system is primarily the scope of this paper. Based on FWI techniques, our online fuzzy expert system can help high school students select a few most relevant colleges to apply. It has an interactive interview process with animated help, and has advisories produced for users at different stages. The output short list of colleges to apply contains five kinds of colleges (highly selective down to non-selective) with possibilities of acceptance in each college according to rules provided. Output also contains a plan of study. The system is also able to find possibility of acceptance in a user chosen college. All data collected is stored, to minimize questions when a student is a repeat user.

SCA is a fuzzy expert agent with hierarchical fuzzy knowledge base using fuzzy logic [1-5]. Users are expected to be able to use SCA as a trusted, involved and smart counselor, who remembers their profile and their past history, advises them through their high school years through all their phases of school, gives them increasingly better and specific nuggets of advice as they make progress through their schools, and guides them toward a selected set of colleges to apply which optimizes their potential. SCA would also be a smart tool to play self-guided “what if” games to explore what specific action one should be taking to have maximum impact on the selection of colleges one can get in and on his/her chances of admission. It employs a natural interactive interview process for a user (i.e., a user can play a game with the SCA to gradually find out rational solutions), and produces advisories at different stages. The college selection with relevant advice is a multi-feature-based complex process of assessment and decision for different domains (major, selectivity, college cluster, and individual college, etc.). The hierarchical fuzzy assessment trees are designed based on hierarchical relationships among features to make a user do detailed assessment on a small group of relevant features at different levels of trees. The leaf nodes in a fuzzy assessment tree represent interview features provided by a user and the internal nodes represent the abstract features generated by interview features. This bottom-up calculation propagates eventually to the root node of the fuzzy assessment tree, and generates a final possibility of college admission (output of the root).

SCA employs a Search Optimizer (SO) module for an effective search over college database periodically. SO is a matchmaker between a user’s profile and search objectives and colleges’ profiles, which can narrow down the search using a small number of screens of user input and well-defined categories. A key problem with today’s search engines is that they are keyword based and their data is open-ended which causes a single search to generate thousands of hits, wasting users time and, worse, frustrating their search effort. For a well-defined domain, data can be categorized precisely, albeit with some effort, and a category/feature-based search can be performed resulting in only few selected hits for each search. Category-based search also allows users to click his/her way to the target information, as opposed to fumbling with correct descriptions/keywords to be employed. Additionally, individual user’s usage pattern can be intelligently tracked, thus enabling users to quickly arrive at their targets with only few clicks by boot-strapping off a related past search. Similar tracking is possible of groups of related users for added functionalities. Such category-based search engines provide efficiency for the users on one hand and allow the database portal owner to effectively organize their backend information on the

other hand. The SCA employs profile and other information gathered from a student to search a database of colleges and output links to the home pages of the colleges found. SO attempts to narrow down the search using not too many screens of user input, but provide access to the current set of colleges found at each stage and the ability to further refine the search using additional screens of student input to the extent possible. Past searches of a student may be employed in the search process.

The major advantage of the hierarchical advisory expert system is that hybrid experience of many experts is used to make a reliable decision, i.e., a student can virtually talk to many experienced experts simultaneously to get robust assessment and stable advice. Similarly, the student can continue to play with the game tree to navigate the partial tree or even the entire tree to find out useful college selection information. These may include (i) top 10 candidate colleges based on the current facts on the game tree and the student’s personal profile, (ii) suggestions on how to improve, for example, SAT score from 1090 to 1200.

#### Basic Contributions:

- We demonstrate how FWI technology is employed in a complex real-life enterprise application.
- We develop new techniques for hierarchical fuzzy calculations of aggregate probabilities for the domain of applications where in basic features can be provided. Without such hierarchies, the execution time of such calculations can quickly rise to exponential complexities. The curse of dimensionality of a regular fuzzy system is a common problem because the number of  $n$ -variable  $m$ -linguistic-value fuzzy rules is  $O(m^n)$ . In our system, the complexity of the new tree structure of a fuzzy rulebase is  $O(m^k)$  where  $k < 6$ , as number of children at any node is bounded.
- SCA can be used in other applications because it is a general framework. Both the search optimizer and the SCA have been designed to allow flexibility and future scalability.

The rest of the paper is organized as follows. Section 2 describes user interface to the system. Section 3 describes the basic architecture and the major functions of SCA. Section 4 presents the new hierarchical fuzzy reasoning method for SCA. Finally, Section 5 gives conclusions.

## 2. User Interface

SCA has a natural interactive interview process for a user. It allows “what if” games, and at any stage, finds out most suitable colleges for students. It produces advisory messages at different stages in order to fully assess the credentials of a student and to guide student’s future development through these advisories. At the heart of this system is an assessment tree, which uses bottom-up fuzzy calculations to generate possibility of admission in various sets of colleges. The output is a short list of colleges to apply containing five kinds of colleges (highly-selective down to non-selective) with possibilities of acceptance in each college according to fuzzy rules provided. All data collected is stored to minimize questions when a student is a repeat user. A new clustering algorithm for personal searches, group searches (peer groups), and all users searches is designed. The user has the ability to refine existing views or new searches to the extent possible and to bookmark the current search at any time. Importantly, a student can virtually interact with many underlying experienced experts behind the agent simultaneously to get robust assessment and sound advice.

### 2.1 Typical Screen Sequence for a First-Time User

A typical first time user comes to the *welcome screen* (Welcome, Basic Introduction, Demonstration, Interview), and goes to a *guided interview screen* (also, *the play screen*). Some may choose quick interview mode, i.e. form filling. Once in *guided interview screen*, the user is queried hierarchically, through repeated updated instances of this screen. The purpose of these sequences of screens is (i) to collect profile features as well as, based on the context, (ii) to give specific advice and (iii) to show user’s possibilities into five categories of colleges. The evaluation employs three phases: (i) Initial Assessment Phase, (ii) Detailed Assessment Phase, and (iii) Major-specific Assessment Phase. User quickly gets familiar with the system and its capabilities, its question hierarchy and phases, and the navigation maps. The effect of each change in an input or output feature during any phase at any level of hierarchy is quickly calculated and the changes in the possibilities of admission is promptly displayed in the next screen. This allows game-like interactivity while user enters additional facts.

All the while, an animated character, “Jill/Jack” interacts with the user explaining the screen components, input questions, output results, and parceling out advices based on the context and input/output values. “Jill/Jack” can also be made to sit on the sidelines, and then the corresponding explanations are available to the user as

“blurbs” hidden within each components, and advices written out along side with the components on the screen.

At any time in the interview process, the user can then ask the system to find colleges to apply, which takes the user to *selected college screen* which shows a few selected optimum colleges to apply, user’s possibilities in each of them, and college specific advices/plan of actions. This screen also shows additional advices/plan of action based on different college categories. User can subsequently choose some of his/her own colleges to be evaluated against user’s profile. The user can go back and forth between *guided interview* and *selected college screens* to change profile and see its effect on college selections and possibilities of acceptances, playing “what if” games.

### 2.2 A Typical Usage Pattern for a Repeat User

A repeat user is prompted with calendar and high-school status (freshman, junior, senior, etc.) based advices/plan of actions, as well as specific questions based on the plan of actions given last time. This allows a repeat user to quickly update his/her profile. The user typically then goes to *guided interview screen* to further update his/her profile and continue with the interview session, or the last “what if” game. Or, the user can edit his/her favorite colleges and see his/her improved chances in the *selected college screen*. Once in the *guided interview/play screen*, a repeat user typically will alter values of one or more input or output features and play “what if” games to determine how his/her chances of getting into different categories of colleges would improve.

To aid in quick navigation through different phases and the hierarchy of these phases, a clickable navigation sub-tree in the *play screen* is always visible (available to the first time users as well). The effect of each change in the input or output feature in any phase at any level of hierarchy is quickly calculated and the changes in the possibilities of admission is promptly displayed in the next screen.

As for a new user, a repeat user would typically go back and forth between *play* and *selected college screens* to see how changes to input/output features alter college selections and possibilities, and what specific future actions he/she must take to get into a specific college or a college category.

### 3. Top Level Technologies

The SCA platform is a Web-based expert system with multiple hierarchical domain-specific knowledge bases allowing many experienced experts to input/update the fuzzy rules. The SCA has unique novel capabilities supported by techniques such as expert system, fuzzy logic, databases, statistics, decision tree, and client-server networking. Here, for clarity, two top-level key techniques are described.

#### 3.1 The Hierarchical Advisory Expert System with Multiple Fuzzy Assessment Trees

Since college selection with relevant advice is a multi-feature-based complex process of assessment and decision for different domains (major, selectivity, college cluster, individual college, etc.), hierarchical fuzzy assessment trees are designed based on hierarchical relationships among features. This allows detailed assessment on a small group of relevant features at each internal node. A fuzzy assessment tree consists of many nodes including leaf nodes representing interview features provided by a user, and high-level internal nodes representing abstract features generated by interview features. Each high-level node (parent node) has a fuzzy rule table that is used to calculate possibility (output of the parent node) based on input features from low-level child nodes. This bottom-up propagation calculation will eventually reach the root node of the fuzzy assessment tree, and generate a final possibility of college admission (output of the root). Corresponding to the four phases of the evaluation process, there are four types of fuzzy assessment trees, each usually five in number corresponding to five categories of colleges (highly selective down to non-selective

1. Five Initial Assessment Trees,
2. Five Detailed Assessment Trees,
3. Five Major Assessment Trees,
4. Multiple College Cluster Assessment Trees, and, possibly in the future, multiple individual college assessment trees.

The key results of the SCA are

1. a short list of colleges with acceptable high possibilities of college admission, and
2. relevant advisory comments and plan of action.

These outputs enable a user not only to understand oneself in possibilities and capabilities in terms of admission to different levels of colleges, but also know what to do to enhance chances to get into the desired colleges in the near future. Thus, the SCA is a real-time

intelligent agent using latest intelligent techniques, advanced databases and efficient networking functions. The SCA (client-server version on the Internet or portable version on a user's local computer) is a short-cut bridge between high school graduation and college admission.

#### 3.2 Implementation Level Architecture

##### 3.2.1 Four Phases of SCA Game

The student interview and evaluation process is a four-phase process. In Phase I, a few features (GPA, SAT and student's self-evaluation scores for strength of his/her curriculum and extra-curricular activities) are employed to find student's *initial assessment* in five categories of colleges (most selective, highly selective, selective, less selective and non-selective – the *five college clusters*). These initial assessments result in five  $PC_1$  values (possibility into college, Phase 1) for the five college clusters. In Phase II, which results in a *detailed assessment* of student's chances in the five college clusters, a student is led through an elaborate and engaging interview process to collect all his/her *basic features* (such as SAT-I score, number of extra-curricular activities engaged in, etc). At each juncture through this interview process, calculations are made by the system and the student receives an update on his or her chances, using a color-coded bar-chart display. The corresponding five possibilities in the five college clusters are collectively referred to as  $PC_2$ 's. In Phase III, student's chances are calculated in his/her desired major (*discipline preference*), resulting in five  $PC_3$ 's. If the discipline preference for the member is missing or undefined, Phase III is skipped. In the last phase, Phase IV, student's *preference profile* is collected and Search Optimizer (SO) platform is employed to filter out 100 or fewer colleges. These colleges are then individually matched against student's features to obtain  $PC_4$ 's (only one value per college). The four sets of  $PC_i$ 's are then combined to obtain student's chances in each of the filtered colleges ( $PC$ s). Finally, a short list of 10 to 20 colleges is produced that the student is recommended to apply in. So, the generated results are very convenient for a user. If the results are not satisfactory, the user may continue to play with SCA.

##### 3.2.2 Rules Trees

Phases I and II employ hierarchical fuzzy rule structures, namely, *Initial Assessment Tree (IA)* and *Detailed Assessment Tree (DA)*. These trees are actually partial orders because a child node could be input to multiple parent nodes. Thus, these structures are directed acyclic graphs on top of the basic features, referred to as trees here for convenience. An example is shown in Figure 1.

Likewise, for Phase III, each major has a *Major Assessment Tree (MA<sub>i</sub>)*, and, for Phase IV, each college has a *College Assessment Tree (CA<sub>j</sub>)*. Thus, the logical structure of all the four phases combined together is a giant partial order (referred to as the *scaTree*) such that SCA is composed of four kinds of sub-trees in the following order: *IA*, *DA*, *MA<sub>i</sub>*, and *CA<sub>j</sub>* (*i* iterates through all major trees and *j* iterates through all college trees). The *scaTree* has unique node numbers for all its nodes, including the *leaves*, which are the basic features of the students, and the *internal nodes*, which are *derived features* of students based on the leaf features of the corresponding sub-tree. *IA* and *DA* trees are unique, but there are several *MA<sub>i</sub>*'s and *CA<sub>j</sub>*'s. Multiple colleges with similar admission profile can have the same *CA<sub>j</sub>*.

### 3.2.3 Miscellaneous Aspects

**Explanation, Advisory and Jill's Phrases:** Each node *X* with children *Y* and *Z* not only has rules table but also carries explanations for features *X*, *Y*, and *Z* and for value gradations of *X*, *Y*, and *Z*. Likewise, there are advisory phrases corresponding to future plan of actions/study (*POS*), for each value gradation of *X*, *Y*, and *Z*. Advisory phrases are displayed once a new value for a feature has been employed to calculate new scores.

Jill has a sound and graphics URL corresponding to each value gradations of all features *X*, *Y*, and *Z* corresponding to explanation phrases in one screen and to advisories in the next screen.

**Base and Session Vectors:** For each member, SCA keeps a folder of student's current basic and derived features into the *base vector*. During a "what if" game, student may change his/her base feature values. The current "what if" play values of student features are kept track of separately in the *session vector*.

**Default Session Vector:** For all subtrees of SCA, namely *IA*, *DA*, *MA<sub>i</sub>* (for all discipline trees *i*), *CA<sub>j</sub>* (for all college trees *j*), a default session vector is initialized manually based on typical values ("Average Joe" values).

**Flexible Framework Design:** The SCA system has been designed as a general framework, allowing administrators to alter the feature trees, the rules, explanations and advisories at each node of the feature trees, and various other tuning parameters at any time in the database without any change in the source code. This will allow them to alter the interview order and context as well as to refine the rules with experience.

## 4. Hierarchical Fuzzy Reasoning Method

### 4.1 Hierarchical Fuzzy Rule bases

Each internal node of *scaTree*, say node *X* with children *Y* and *Z*, has a rules table, with input columns corresponding to *Y* and *Z*, and five output columns corresponding to the strengths of the derived feature *X* for the five college clusters. If child *Y* has five range of values and *Z* has six range of values, then the rules table has 30 rows, one for each combination of values of *Y* and *Z*. Logically, therefore, the rules table is five different tables for five college clusters, one corresponding to each of its output column (Exceptions are tables for *CA<sub>j</sub>*'s which have just one output column).

Each value range (*value gradation*) for an input column, say *Y*, has a typical value (*center value*, *C*), a low value (*L*) and a right, upper value (*R*). Neighboring ranges may not be disjoint, so *R* for a range may be greater than *L* for the following range. Overlapping ranges lie at the core of fuzziness. In the above example, therefore, a specific pair of values of *X* and *Y* may correspond to up to four rows in the rules table, and therefore "fire" the corresponding rules. Thus, the value calculated for node *X* is robust, employing fuzzy logic.

### 4.2 Fuzzy Knowledge Base

The fuzzy knowledge base consists of a set of rules and a set of membership functions. Experts often incorporate in their reasoning process qualitative concepts such as high, low, hot, slow, etc. The inherent vagueness of these terms (called *linguistic values*) is naturally implemented by fuzzy logic sets, which in turn are characterized by their membership functions [1]. Also, this same expert not only often think in IF-THEN patterns when solving problem but also represent and employ expertise with fuzzy concepts. As a result, the rules in the fuzzy rulebase are usually of a form similar to the following:

IF *X* is low and *Y* is high then *Z* is medium

where *X* and *Y* are inputs variables, *Z* is an output variable, and low, high, and medium are membership functions (fuzzy subsets) on *X*, *Y*, and *Z* respectively. *X*, *Y*, and *Z* are the linguistic variables.

The antecedent (the rule's premise) describes to what degree the rule applies, while the conclusion (the rule's consequent) assigns a membership function to each one or more output variables. More than one conclusion per rule is allowed. The set of rules in the fuzzy rulebase is known as the *rule base*.

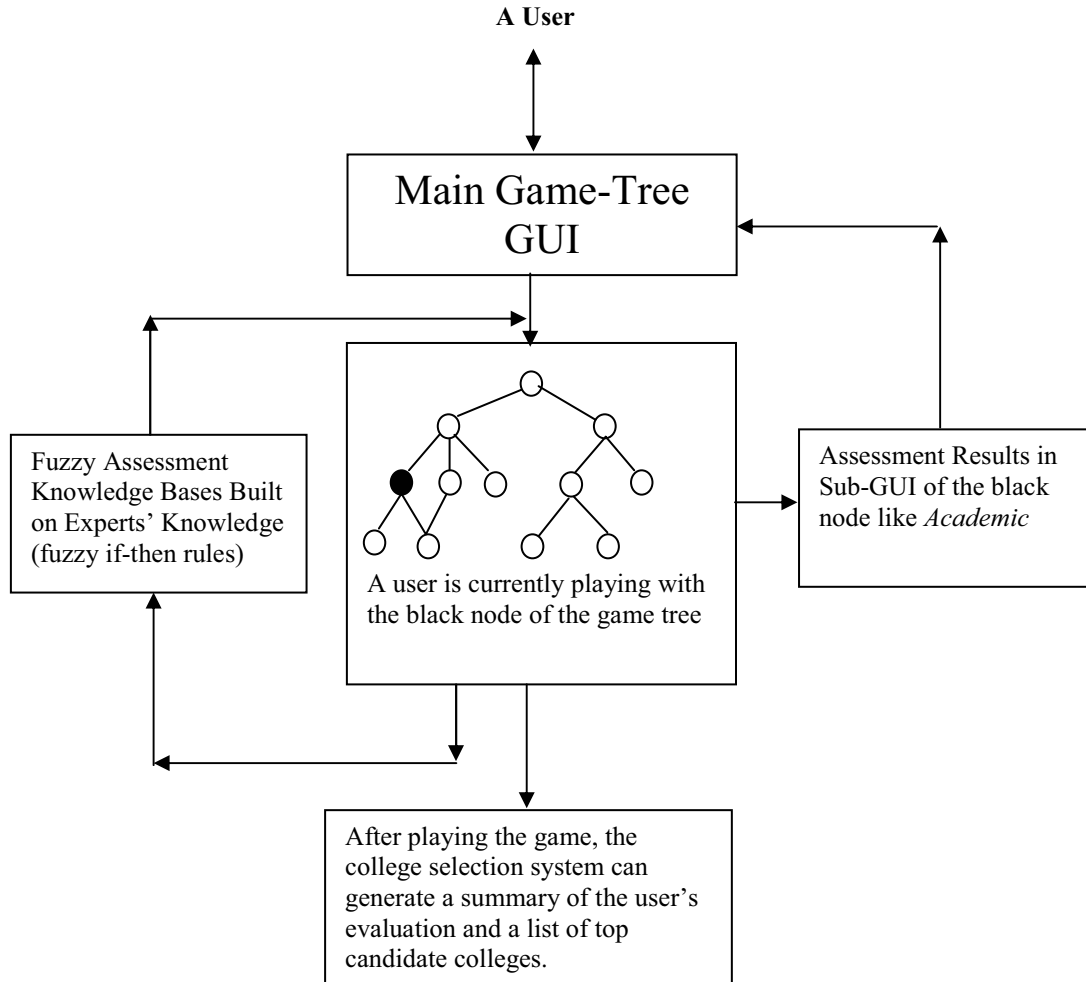


Figure 1: The College Selection Game Tree

### 4.3 Fuzzy Inference Engine

This is the heart of the Fuzzy Expert Agent (FEA). It is also known as fuzzy reasoning. It is an inference procedure that derives conclusions from a set of fuzzy if-then rules and known facts. In a FEA, the reasoning procedure involves three or four primary processes: fuzzification, fuzzy inference, composition, and defuzzification. Some people consider composition and defuzzification as one process.

**Fuzzification:** Fuzzification operations are used to combine a real-time input value (e.g., temperature and speed) with stored membership function information to produce fuzzy input values. In other words, under fuzzification, the membership functions defined on the input variables are applied to their actual values to determine the degree of truth for each rule premise.

**Fuzzy Inference:** Fuzzy Inference attempts to relate the fuzzified input facts to the premise patterns of fuzzy rules. Under inference, the truth-value for the premise of each rule is computed, and applied to the conclusion part of each rule. This results in one fuzzy subset to be assigned to each output variable for each rule. Usually only MIN or PRODUCT is used as inference methods or inference rules [4].

**Composition:** Composition combines all fuzzy outputs into a specific composite outcome. Under composition, all of the fuzzy subsets assigned to each output variable are joined together to form a single fuzzy subset for each output variable. The MAX or SUM are generally used to do the combining.

**Defuzzification:** The centroid defuzzification method (COA) calculates the weighted average of a fuzzy set [4]. The result of applying COA defuzzification to a

fuzzy conclusion “Y is A” can be expressed by the formula:

$$y = \frac{\int u(y_i) \times y_i dy}{\int u(y_i) dy}$$

## 5. Conclusions

The Web-based Smart College Agent (SCA) with hierarchical fuzzy knowledge base and flexible fuzzy reasoning methods can help users search college information on the Internet efficiently and make good decisions effectively by using complete fuzzy assessment trees. The SCA has a game-playing-oriented GUI for making users enjoy the “games” controlled by SCA. SCA has a natural interactive interview process for a user, and produces advisory at different stages in order to fully assess the credentials of a student and find out most suitable colleges for him/her. Importantly, it is a general framework that can be used in other applications such as e-Commerce, expert assessment and self-evaluation. In the future, other computational Web intelligence techniques [6, 7] can be also applied to SCA to make them have online learning ability.

## 5. References

[1]. <http://www-bisc.cs.berkeley.edu/>. M. Detyniecki, "Fuzzy Decision Trees: History and Challenges,"

*Berkeley Initiative in Soft Computing (BISC) Seminar*, UC Berkeley, December 2001.

[2]. <http://www.abo.fi/~rfuller/ifsa.html>.

[3] <http://morden.csee.usf.edu/Nafipsf/>. M. Shimoda, [A natural interpretation of fuzzy sets and fuzzy relations](#), *Fuzzy Sets and Systems*, 128 (2002)

[4] Yen J. and Langari R. (1999). *Fuzzy Logic: Intelligence, Control, and Information*, Prentice Hall.

[5] Zadeh L.A. (1965). Fuzzy Sets. *Information and Control*, 8, 338-353.

[6] Zhang Y.-Q., and Lin T.Y. (2002). Computational Web Intelligence (CWI): *Synergy of Computational Intelligence and Web Technology*. *Proc. of FUZZ-IEEE2002 of World Congress of Computational Intelligence 2002: Special Session on Computational Web Intelligence*, pp. 1104-1107.

[7] Zhang Y.-Q. and Kandel A. (1998). Compensatory Genetic Fuzzy Neural Networks and Their Applications, *Series in Machine Perception Artificial Intelligence*, Volume 30, World Scientific.

[8] A. Parvatiyar, S.K. Prasad, R. Sunderraman and Y. Zhang, “Smart Advisor and Search Optimizer: Web-based Applications of Fuzzy Rules, Intelligence Systems and Hierarchical Clustering for Relational Decisions,” *the Sixth Research Conference on Relational Marketing and CRM*, June 6-12, 2002