URBANO: A Tour-Guide Robot Learning to Make Better Speeches

José Javier Rainer, Ramon Galán
Intelligent Control Group. Universidad Politécnica de Madrid
C/ José Gutiérrez de Abascal 2. 28006 Madrid SPAIN
e-mail: javier.rainer@upm.es; ramon.galan@upm.es

***Abstract*—Thanks to the numerous attempts that are being made to develop autonomous robots, increasingly intelligent and cognitive skills are allowed. This paper proposes an automatic presentation generator for a robot guide, which is considered one more cognitive skill. The presentations are made up of groups of paragraphs. The selection of the best paragraphs is based on a semantic understanding of the characteristics of the paragraphs, on the restrictions defined for the presentation and by the quality criteria appropriate for a public presentation. This work is part of the ROBONAUTA project of the Intelligent Control Research Group at the Universidad Politécnica de Madrid to create "awareness" in a robot guide. The software developed in the project has been verified on the tour-guide robot Urbano. The most important aspect of this proposal is that the design uses learning as the means to optimize the quality of the presentations. To achieve this goal, the system has to perform the optimized decision making, in different phases. The modeling of the quality index of the presentation is made using fuzzy logic and it represents the *beliefs* of the robot about what is good, bad, or indifferent about a presentation. This fuzzy system is used to select the most appropriate group of paragraphs for a presentation. The *beliefs* of the robot continue to evolving in order to coincide with the opinions of the public. It uses a genetic algorithm for the evolution of the rules. With this tool, the tour guide-robot shows the presentation, which satisfies the objectives and restrictions, and automatically it identifies the best paragraphs in order to find the most suitable set of contents for every public profile.**

Keywords- Cognitive systems; learning; autonomous robot; fuzzy systems; decision making.

I. INTRODUCTION

Autonomous robots are intelligent machines capable of performing tasks in the world by themselves, without explicit human control over their actions [1].

Within the development of multiple applications for a mobile robot, probably one of the first real world applications of indoor service robots has been mobile robots serving as tour guides in museums or exhibitions. We have developed our own interactive mobile robot called Urbano specially designed to be a tour guide in exhibitions [2]. The basic characteristics of Urbano are described in Section 3.

Our goal has been to create an automatic presentation generator that allows the flexible and dynamic display of information depending on the distinct kinds of audiences and other parameters that characterize the presentation. The automatic selection of contents for composing sophisticated presentations is a non-trivial task. If the aims of the presentation, preference and interest for a particular subject have to be taken into account, it becomes even more complex.

The knowledge is based on an ontology of domain-specific concept words. Ontologies have been known in computer science as consensual models of domains of discourse, usually implemented as formal definitions of the relevant conceptual entities [3].

Some of the most recent works about automatic generation are: [4][5][6][7][8]. These works propose different architectures and methodologies than those presented here. What is original is the introduction of fuzzy logic in the pruning of the resulting tree and in the quality index of the presentations.

This paper is focused on the automatic generation of presentations by a robotic system. The MINERVA robot from Carnegie Mellon University was one of the first service robots for guiding tours. Others were Rhino, from Bonn University, or Xavier, built to participate in AAAI Robotic Competition in 1993.

Aspects concerning the grounding of the symbols used by the robot, e.g., paragraphs in the presentation or the learning of new concepts – have been circumvented for problems of space. In [9], the authors explore the problem of learning and the symbol-grounding problem, and propose a systemic and integrative approach both problems.

II. AUTOMATIC TEXT GENERATION

The history of natural language processing is characterized by the influence of artificial intelligent [10][11]. A natural language generator generally has access to a wide knowledge field from which it must select information to present to the users in various formats. Generating text is, therefore, a decision-making problem with multiple restrictions: knowledge, available linguistic tools, the objectives of the user to whom the text will be directed, the situation, and past discourse. It deals with identifying the factors involved in this process and determining the factors that intervene in this process and its consequences [11][12][13].

An additional contribution of our work is to provide a dynamic framework that allows us to take into account each possible scenario. On the other hand, there can be no single general-purpose presentation format for all users, because

Copyright (c) IARIA, 2010 ISBN: 978-1-61208-108-3

each user differs in all aspects of interests and expertise levels, and in the devices used to visualize the requested information. These aspects are used as restrictions.

As a result, to prepare a presentation, the candidate items can be multiple, depending on knowledge server. The items selected in the generation process contain information about the theme of the presentation, but must as well include the items that are semantically related to the presentation. Thus, for example, a discussion of the painter Velazquez will also speak about the Italian painter Caravaggio due to his influence on Velazquez.

In our case a presentation generator is considered as a Cognitive Skill. It is assumed that skill development, e.g., giving presentations progressively more suitable to the public being addressed – is a fundamental architectural epiphenomenon. Rather than viewing a presentation as a mere form of communication, the focus here is on leveraging it as a means of expanding socio-communicative skills.

Several research projects have been undertaken to develop software tools for generating narratives, histories or presentations, and they have described many characteristics for generating presentations, but the quality criteria vary [14][15].

The decision rules to establish the fact that a particular swap in the presentation strategy is useful and required are not clear yet. An interesting situation arises when the data changes and the environment are dynamic.

A. Quality criteria of a presentation

The quality of a presentation is defined by the different aspects that it characterizes, referred to in [16], as: nature, purpose, duration, and number of participants; for other authors, such aspects as connecting with the audience, the interests of the participants, the first and last minute, changes of rhythm, and "being natural and having fun" are very important. All the aspects have in common the difficulty of being effectively quantified for a computer program.

The prototype as developed has proposed the following *quality criteria*:

- Differences in time used and the anticipated time
- Time dedicated to the theme
- Time dedicated to entertainment
- Difficulty in understanding the theme
- Interest in the theme
- Time dedicated to related information (anecdotal)
- Time for interaction with the public
- Original information
- Non-conventional focus

These criteria are also used to obtain the assessment of the public.

III. URBANO

This Section describes the Urbano robot system, its hardware software and the experience we have obtained through its development and use until its actual mature stage.

A. URBANO, an interactive Mobile Tour-Guide Robot

This Section doesn´t pretend to be an exhaustive technical description of algorithms, mathematical or implementation detail, but just an overview of the system.

Urbano robot is a B21r platform from iRobot, equipped with a four wheeled synchrodrive locomotion system, a SICK LMS200 laser scanner mounted horizontally in the top used for navigation and SLAM, and a mechatronic face and a robotic arm used to express emotions as happiness, sadness, surprise or anger.

The robot is also equipped with two sonar rings and one infrared ring, which allows detecting obstacle at different heights that can be used for obstacle avoidance and safety. The platform has also two onboard PCs and one touch screen.

The software is structured in several executable modules to allow a decoupled development by several teams of programmers, and they are connected via TCP/IP. Most of these executables are conceived as servers or services providers, as the face control, the arm control, the navigation systems voice synthesis and recognition, and the web server. The client-server paradigm is used, being the only client a central module that we call the Urbano Kernel. This kernel is the responsible of managing the whole system [2].

URBANO robot has a technology based on distributed application software. The recent version is an agent based on architecture that uses a specific CORBA approach as an integration tool. The robot has many functions: speaks, listens, navigates through the environment, moves his arm, responses to stimuli that affect its feelings.

Figure 1. Urbano Tour-Guide Robot.

B. URBANOntology

The knowledge server consists of a Java application developed using the libraries of Protégé-OWL API. The tool is capable of reading and editing files in ".owl" format where the knowledge is stored in the form of ontologies and the management of the information from the kernel is made by means of messages that codify the request of specific information, and the reply is obtained from the server or the introduction of new data.

Copyright (c) IARIA, 2010 ISBN: 978-1-61208-108-3

The functions of the knowledge server are: loading and saving ontologies; creating, renaming, and deleting classes or instances; displaying properties of a class; showing subclasses or superclasses; showing or entering the value of a property; integrating one ontology into another; handling queries.

IV. APG AGENT

APG agent software has been developed to be integrated in the architecture based on the agents that constitute the software of the Urbano robot. The developed computer application is activated on receiving a request from a user that selects a file that contains the pattern to be developed.

APG will request all the information from the knowledge server agent, using restrictions that it needs to generate a file with the best presentation to be used by the system in the next performance by the robot.

Figure 2 shows this flowchart. It must be mentioned that presentations have been stored with quality indices (QI) for each of them, and in some cases, when faced with a request very similar to a presentation already performed with high quality indices; this presentation will be used without needing to generate a new one.

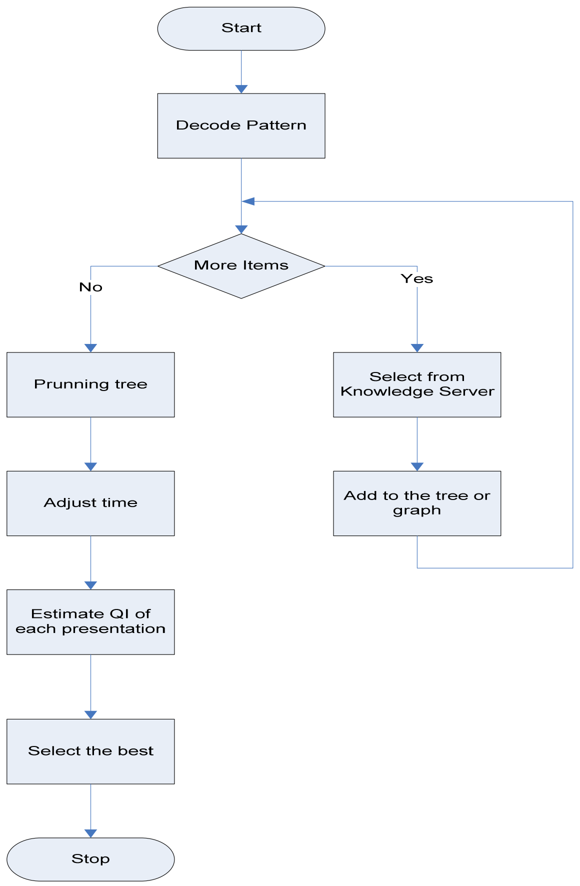

Figure 2. Flowchart APG

V. PARAGRAPHS

The paragraph is used by the robot as the minimum element of expression. It is assumed that the paragraphs have a size such that coincidences or references between them lack meaning, and that they express a meaningful content. The following paragraph introduces the painting “Las Meninas” by Diego Velazquez:

“A portrait of the infanta Margarita, daughter of Felipe IV (1605-1665), surrounded by her servants or ‘family’ in a hall of Madrid’s Alcázar Palace.

This, the most famous of Velasquez’s works, offers a complex composition built with admirable skill in the use of perspective, the depiction of light, and the representation of atmosphere.

There have been innumerable interpretations of this subject and later references to it. The most numerous emphasize a defense of the nobility of painting versus craft. Velasquez portrays himself, painting the painting itself, on the left of the canvas, thus affirming the supremacy of the art of painting. The infanta Margarita (1651-1673), wears white and appears in the center of the composition, surrounded by her ladies in waiting, the “meninas” María Agustina de Sarmiento and Isabel de Velasco, along with two court buffoons, María Bárbola and Nicolasito Pertusato, and a mastiff. Behind her, the duenna Marcela de Ulloa converses with the quartermaster, José Nieto, who is in the doorway.

The King and Queen, Felipe IV and Maria de Austria (1634-1696) are reflected in the mirror at the back of the room, leading to series of extraordinarily complex spatial relations.”

For each of the paragraphs there also exists a script that specifies the facial expressions and arm movements of the robot, the tone of voice and type of voice to use, the position in the floor plan, and various details for interaction with the public.

But in the ontology, each paragraph is “related” to the categories, which it belongs (e.g., in the previous paragraph: Is_a Painting, Is_a Description, Is_a Adult Level, Painted_by Diego Velazquez, etc. These relationships are used for the selection of the paragraph).

VI. PATTERNS AND TREES

The pattern, defined by the user, signals the elements that should form part of the presentation. For each item, these elements are established: its identification, its priority, its numerical order, its reference to the theme within available knowledge. There could be a limitation of time and a very large patter, in that case the priority index indicates the most important content to be included, but there could be the opposite case, that there is time left, then the system incorporates content that it is not initially considered.

It uses XML as the language to represent the patterns, which guarantees an easy use with different tools and programming languages. XML has emerged as a de facto standard for encoding and sharing data between various applications. XML is also useful for structured information

Copyright (c) IARIA, 2010 ISBN: 978-1-61208-108-3

management, including information contained in knowledge server [18]. Our proposal is to design a software tool that helps the user to create and maintain the patterns.

APG requests from the knowledge server the paragraphs available for each theme, identified by its reference. The knowledge server will submit one or more paragraphs, because of a same concept can have several versions. Thus the pattern indicated in Figure 3 shows that first item of the presentation will be “museum_presentation” identified on the knowledge server as “Greetings” and it could involve, for example, three possible paragraphs, which will be included in the tree of possibilities, as is shown in Figure 4.

Some global restrictions including the level of audience will be used to prune the tree, eliminating the paragraphs that do not correspond to the requested level.

Three typical alternative heuristic searches have been tested to trim the tree. The first uses “brute force” to generate all the possible combinations and to group all the numeric values of the “quality criteria” of the paragraphs that form the presentation, and then, using a set of fuzzy rules, estimates the quality index. It selects the presentation with the highest index. See Figure 5.

The second alternative uses “best-first search” so that as it goes along it takes the option that partially presents the best index. This alternative is without a doubt the fastest, but it cannot guarantee the selection of the best option.

The third alternative modifies the previous one so that it generates “backtracking” if the quality index falls below a minimum.

```
<pattern>
    <id>children_visit_museum</id>
    <date_creation> … </date_creation>
    <date_lastused>…</date_ lastused>
    <restrictions>
        <age><10</age>
        <technical_level>low</technical_level>
        <time>60</time>
    </restrictions>
    <item>
        <item_id>Museum_ Presentation</item_id>
        <item_order>1</item_order>
        <item_priority>10</item_priority>
        <item_data>Greetings</item_data>
    </item>
    <item>
        <item_id>Tour_Guide</item_id>
        <item_order>2</item_order>
        <item_priority>20</item_priority>
        <item_data>Tour_Guide</item_data>
    </item>
    <item>
    <item_id>Painting_presentation</item_id>
        <item_order>3</item_order>
        <item_priority>40</item_priority>
        <item_content>
            <item_data_id>key_picture</item_data_id>
            <subitem> title</subitem>
            <subitem>date</subitem>
                <subitem_multiple>description</subitem_multiple>
            <subitem_multiple>period</subitem_multiple>
        </item_content>
    </item>
</pattern>
```

Figure 3. XML Pattern Example

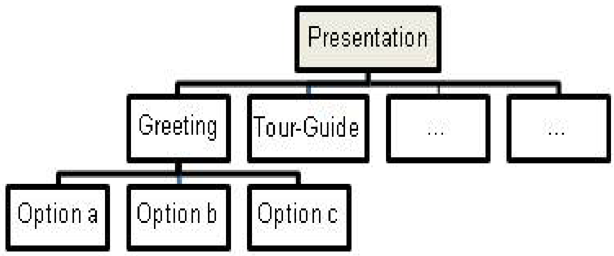

Figure 4. Tree Data Structure

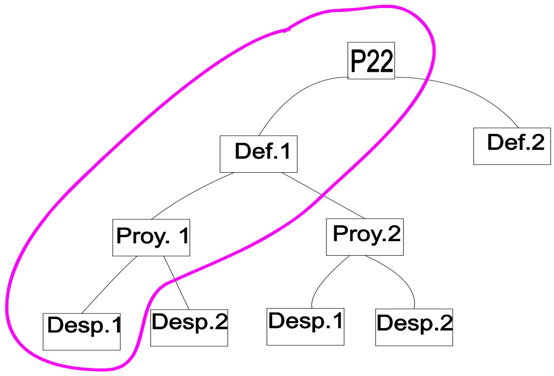

Figure 5. Different options

The presentation generated in this way analyzes the estimated time for its execution, and if this is greater than anticipated, it eliminates the paragraphs with the least necessary priority. On the other hand, if there is enough time, it includes some socially-oriented paragraphs such as jokes or comments about sports, politics, or local events.

VII. Decision-Making

Decision-making is a part of the paradigm proposed by Zadeh [19] that has been currently examined in [20]. In a dynamic scenario as ours, and because of the nature of the information that the system will handle, proper tools are needed to provide the intelligence for decision-making and supervision.

Decision-making is the cognitive process of selecting a course of action from multiple alternatives. Fuzzy set approaches to decision-making are usually most appropriate when human evaluations and the modeling of human knowledge are needed [21].

The proposed solution uses fuzzy rules to prune items to submit from the tree. It will use a variable that indicates the likelihood of inclusion in the submission of a particular content. The fuzzy system will generate these values.

The fuzzy rules enable more flexibility. These rules will be adjusted and expanded.

All information available at the moment about the quality criteria and its influence on the quality index is stored in the ontology of the knowledge server.

The semantic network will indicate that the influence of time dedicated to the theme, expressed in a percentage, is

VERY favorable if the time is VERY_HIGH but LOW favorable if the time is HIGH or NORMAL, and VERY unfavorable in any other case. This relationship is defined by belonging to specified categories. In the developed prototype, the fuzzy rules used to calculate the quality index are obtained by consulting the categories that belong to a criteria in the knowledge server.

Five linguistic terms are defined: VERY_HIGH (VH), HIGH (H), NORMAL (N), LOW (L), VERY_LOW (VL). The fuzzyfication phase uses the function of membership to initially equidistant triangles, but in the learning phase their centers can vary. The exit variable quality_index is also modeled with five terms and triangular functions. The technique of centroid method is used in the defuzzyfication phase.

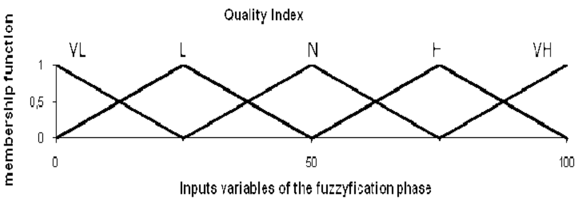

Figure 6. Inputs variables of the fuzzyfication phase

An interesting aspect is the partial quality estimation for the presentation. When considering three paragraphs (#p1,#p2,#p3), for example, it may happen that some criteria have unrepresentative values; if, for example, #p1,#p2,#p3 correspond to paragraphs of technical description, the entertainment criteria will be null and the presentation will have a low estimation. In the tests made using pruning the best-first search, presentations are obtained with very poor estimations, and to compensate for this effect, an option of the best-first was introduced, but rejecting presentations with a low index.

If the categories of information expressed in the pattern are very generic, the number of possible presentations increases enormously, but it permits the robot to generate higher quality presentations.

VIII. Learning Phase

The most important characteristic of the proposal is the ability of the robot to learn. Initially it is thought that the robot will have a small number of quality criteria available to evaluate some presentations as good and others as bad, corresponding to the minimum level of education for a professional guide, in order to guarantee a minimum level of quality in its presentations.

A simple questionnaire has been designed that the public can fill out after attending a presentation by the robot. It asks for an evaluation of each quality criteria known at the time, indicating whether the robot should spend more or less time on each item, and a percentage evaluation of what the public considers valuable in the presentation. The Table I shows an example.

A proper statistical treatment of the questionnaires, eliminating the extremes and requiring a minimal quantity of data is performed.

TABLE I. Quality Criteria

Quality Criteria	**Should be**
Time spent on the theme	**-**
Time spent on entertainment	**+**
…	
New criteria to bear in mind	**%**
Answers to questions	25
Global evaluation	**%**
	60

A genetic algorithm is used, an adjustment of linguistic terms and the membership functions, will permit the quality index to be the closest to the average expressed by the public. To get a greater accuracy, the genetic algorithm is simultaneously used over several presentations as an attempt to offset the local minimals for a presentation.

Learning is realized when new criteria of quality are incorporated in the paragraphs.

The genetic algorithm realizes a readjustment of the rules when it produces a disparity between public opinion and the IC. It utilizes a genetic algorithm for the evolution of the rules

The + and – indications are used to eliminate individual cases generated by the genetic algorithm in which, while still generating a correct quality index, the evaluation of the criteria runs contrary to public opinion.

In the tests carried out, it was shown that the “beliefs of the robot” about the quality indices converge toward public opinion, and as a result generate presentations that are evaluated more favorably.

IX. Conclusion

As a conclusion, special attention has been given to a mechanism for automatic generation of presentations, analysis of search algorithms, learning phase and optimization of fuzzy logic rules, taking into account the intrinsic difficulty of natural language processing and automatic generation.

The generator is a dynamic system where knowledge will increase and, therefore, it will do the quality of the presentations. The robot will become increasingly better when making tour-guided visits. It also adds the ability to gesticulate while is conducting the presentation: it specifies the facial expressions and arm movements of the robot, the tone and type of voice to use, the position in the floor plan, and several other details for interaction with the public.

This development will allow tour-guide robots to offer more affective learning and a dynamic tour-guide visit, because the public generally has become more sophisticated and, also, its expectations and demands. It provides a better use of knowledge management. At the same time, advances in a new visitor-oriented approach, progress towards the creation of modular and scalable scenarios.

It presents a novel approach to the use of a computer ontology to represent the corpus that the robot works with, and the quality criteria for a presentation.

Methodology designed for the automatic generation of presentations can be used for other applications related with decision making for autonomous robots. In the Urbano project design; there is a scheduler agent that selects the best task between different tasks, which the system has to do in a day. The criteria used for selecting the adequate task in short-term plan can be revised to optimize the long-term objective. A happiness model, as long-term objective, and a decision-taking mechanism, as short-term planner, was modeled for Urbano. Both use fuzzy logic and are adjusted by genetic algorithms that use the public opinion to learn to be a good tour-guide robot.

These systems have special importance to develop learning support for environments that require greater motivation and commitment, such as classrooms and workshops for students with special educational needs.

ACKNOWLEDGMENT

This work has been partially funded by Spanish Ministry of Science and Technology (ROBINT: DPI2004-07908-C02-01 and ROBONAUTA: DPI-2007-66846-C02-01)

REFERENCES

[1] G. Bekey, Autonomous Robots: from biological inspiration to implementation and control, MIT Press books, 2005.

[2] D. Rodríguez-Losada, F. Matia, R. Galán, M. Hernando, J. M. Montero, and J. M. Lucas, Urbano, an Interactive Mobile Tour-Guide Robot. Advances in Service Robotics. Ed. H. Seok. In-Teh, 2008, pp 229-252.

[3] M. Uschold and M. Grüninger, “Ontologies: Principles, Methods, and Applications,” Knowledge Eng. Rev., vol. 11, 1996, pp. 93–155.

[4] P. Barnaghi and S. Kareem, “Ontology- Based Multimedia Presentation Generation,” IEEE TenCon, 2005.

[5] S. Bocconi, “Automatic Presentation Generation for Scholarly Hypermedia,” 1st International Workshop on Scholarly Hypertext at the fourteenth conference on Hypertext and Hypermedia (HyperText 2003), August 30, 2003.

[6] A. Celentano and O. Gaggi, “Schema Modelling for Automatic Generation of Multimedia Presentations,” SEKE, 02 Ischia Italy, 2002.

[7] F. Frasincar and G. Hoube, “XML-Based Automatic Web Presentation Generation,” Proc. WebNet World Conference on Www and Internet, 2001.

[8] S. Little, J. Geurts and J. Hunter, “Dynamic Generation of Intelligent Multimedia Presentations though Semantic Inferencing,” The Sixth European Conference on Research and Advance Technology for Digital Libraries, Rome, Italy EDCL, 2002.

[9] J. Gómez, R. Sanz and R. Galán, “Learning in Technical Systems: a Sign based Approach,” In proceedings 1st International Workshop on Cognition for Technical Systems, Munich, 6- 7 October 2008.

[10] R. Cole , J. Mariani , H. Uszkoreit, A. Zaenen, and V. Zue, Survey of the State of the Art in Human Language Technology. Cambridge: Cambridge University Press, 1996.

[11] K. Sparck Jones, Natural languages processing: a historical review. University of Cambridge, 2001.

[12] J. Allen, Natural Language Understanding, Redwood City, Ca.: Benjamin /Cummings, 2ª ed., 1995.

[13] J. Allen, “AI Growing Up. The Changes and Opportunities”, AI Magazine, 19 (4), 1998.

[14] D. Jurafsky, and J. Martin, Speech and Language Processing. Prentice-Hall, 2000.

[15] K. Brooks, Metalinear Cinematic Narrative: Theory, Process and Tool. Phd Thesis, MIT, 1999.

[16] M. Cavazza, F. Charles, and S. J. Mead, “Character-based interactive storytelling,” IEEE Intelligent Systems: Special Issue on AI in Interactive Entertainment, 17(4):17–24, 2002.

[17] F. Laure, Técnicas de Presentación. Cecsa, 2002.

[18] M. Y. Maarouf and S. M. Chung, “XML Integrated Environment for Service-Oriented Data Management,” 20th IEEE International Conference on Tools with Artificial Intelligence, 2008.

[19] L. A. Zadeh, “Fuzzy sets,” Information and Control, 8, 338–353, 1965.

[20] L. A. Zadeh, “Is there a need for fuzzy logic?“ Information Sciences, 178(13), 2751–2779, 2008.

[21] C. Kahraman, “Fuzzy set applications in industrial engineering”, Information Science, Volume 177, Issue 7, Pages 1531-15321, 2007.