Goal Orientated Conversational Agents: Applications to Benefit Society

Keeley Crockett, James O'Shea, and Zuhair Bandar

The Intelligent Systems Group, School of Computing, Mathematics & Digital Technology,
Manchester Metropolitan University, Chester Street, Manchester, M1 5GD, UK
{K.Crockett,J.D.OShea,Z.Bandar}@mmu.ac.uk

Abstract. Goal-orientated Conversational Agents are a specific family of conversational agents that are designed to converse with humans through the use of natural language dialogue to achieve a specific task. Traditionally, they utilise pattern matching algorithms to capture the values of specific attributes through their values through dialogue interaction with a user. This is achieved through the use of scripts which contain sets of rules about the domain and a knowledge base to guide the conversation towards achieving a specific goal. Such systems are ideal for providing clear and consistent advice 24 hours a day in many different scenarios, including advising employees about their organisations policies and procedures, guiding a user through buying a suitable product, and tutoring a student to understand a learning objective. This paper presents an overview of a methodology for constructing goal orientated conversational agents. Three case studies which employ this methodology are introduced and evaluated.

Keywords: Goal-orientated conversational agents, pattern matching.

1 Introduction

A conversational agent (CA) is an agent which uses natural language dialogue to communicate with users [1]. The CA will also have the ability to reason and pursue a course of action based on its interactions with humans and other agents. An automated and interactive conversational agent system should be able to provide anonymous and 24-hour access to humans in a variety of applications e.g. customer self service on the web, internal advice on an organisation's policies and procedures, and would allow users to be able to ask questions in natural language. At the same time, the information / advice given by the CA would always be consistent, appropriate and valid whilst the agent can be designed to exhibit sympathetic or compassionate behaviour to a particular circumstance. A further strength is that CA's can be tailored to behave in a way that reflects an organization's or businesses culture and to have distinctive personalities.

A Goal-Oriented CA (GO-CA) is a type of conversational agent which has a deep strategic purpose which enables it to direct a conversation to achieve a goal [2]. Thus the GO-CA [2, 3] may spend more time leading the conversation and asking questions

than the human participant which makes it different from systems such as chatterbots and traditional conversational agents. Chatterbots [4] attempt to prolong social conversational with a human for as long as possible generally using pointless chat [2]. Traditional conversational agents either focus on the embodied component (e.g. graphical representation of the agents face) with simple dialogue or are text based and require complex scripting procedures to appear to have intelligence in answering the user's enquiry. This paper first introduces a brief history of conversational agents and in section 3 a methodology for constructing GO-CA's is introduced. Sections 4, 5 and 6 describe three diverse applications of GO-CA which illustrate the diverse application of the GO-CA methodology. Finally section 7 highlights some of the main issues in developing CA's and areas of future research.

2   Milestones in Conversational Agent Development

The best known early CA was Eliza [6]. ELIZA's main trick was to use questions to draw a conversation out of the user. However the main criticism of ELIZA was the program's lack of an internal world model that could influence and track conversation [6]. ELIZA was followed by PARRY [7] which simulated paranoid behaviour effectively in a blind, Turing like, test by doctors. This took advantage of the fact that unusual responses from PARRY would be attributed to paranoia. PARRY was followed by a number of CA's which utilised pattern matching techniques based on Infochat [8] which implemented an interpreter for a language known as Pattern Script. The most recent innovation in pattern matching is symbolic reduction which is used in A.L.I.C.E. [9]. Symbolic reduction allows a rule to remove part of the user utterance and re-submit the remainder to the CA for further analysis.

A significant proportion of CA research has been dedicated towards embodied agents where expressions and gestures contribute as much as the natural language dialogue [10]. The purpose is to make a computer application appear more human like and become more engaging with the user. While substantial amounts of work [11, 12] have gone into evaluating such interfaces in terms of features such as visual appearance, expressiveness, personality, presence, role and initiative, less attention has been applied to the evaluation of the actual conversation.

CA's are also being used to allow non-programmers to query databases [13,14]. Owda et al [13] proposes the use of goal oriented conversation to provide the natural language interface and helps disambiguate the user's queries, by utilizing dialogue interaction. CA's have been successfully used within Conversational Intelligent Tutoring Systems (CITS) [15]. CITS are computer-based educational systems which employ intelligent technologies to provide individualised instruction in natural language and help students by providing personalised tutoring at a time and a pace to suit the individual. Case study three will describe CITS in more detail. A number of companies have developed CA's to either sell products or provide customer support. VirtuOz, use conversational agents for customer support such as in online sales, advice and recommendations [16].

3   Constructing Goal Orientated Conversational Agents (GO-CA)

This section describes an architecture developed by the Intelligent Systems Group at MMU for constructing GO-CAs [2]. The methodology is designed so the user can approach the GO-CA with a problem or a request for information which can then lead to interactive dialogue between the user and the GO-CA until the user's goal is achieved. This is achieved in current implementations using a rule-based system, which contains a model of the problem domain that is expressed in terms of a set of attributes. Through the process of dialogue, appropriate attributes are captured (through pattern matching algorithms) to model the particular problem experienced by the user and identify the appropriate solution. A further strength of a GO-CA is that it will engage in extended dialogue, during the course of which it will appear to have mental states that are directed at states of affairs in the world e.g. beliefs, desires, hopes, perception, intention etc [2].The GO-CA is a mixed-initiative system (from time to time either the human or the agent may take control of the conversation) [2]. However, due to the nature of the GO-CA, if the human diverts the conversation from the goal, the GO-CA will always attempt to get the conversational back on track using the rule base and knowledge obtained from attribute values captured during the conversation. Figure 1 shows the generic modular architecture for a typical GO-CA [3]. This is intended to take on challenging real-world applications in which the user may present adversarial, disruptive or deceptive behavior at times during the conversation [2].
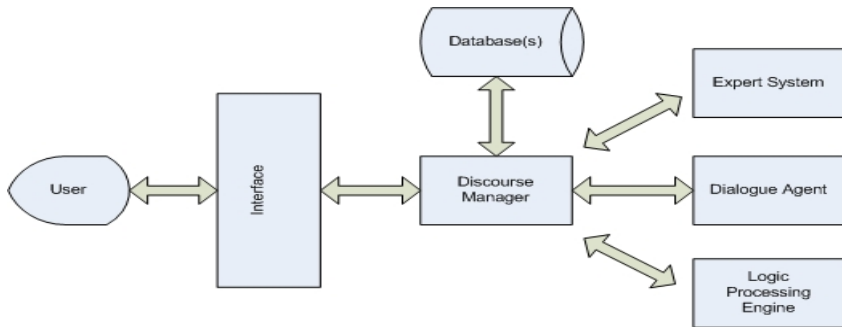

Fig. 1. GO-CA Architecture

Each component within the architecture will now be briefly described:
- The *discourse manager* is the central component of the GO-CA and provides communication with the problem specific expert system (rule base), dynamic graphical user interface (GUI) and the Dialogue Agent (DA). When events take place in the GUI e.g. the user enters dialogue, the discourse manager requests a response from the expert system or the DA, and then instructs the GUI to update its display accordingly.
- The *expert system* contains a rule base comprising of the domain knowledge. Through dialogue with the DA, the rule base will gather information about the user's circumstances and will inform the discourse manager which attributes and its associated value needs to be captured in order to progress.

- The *graphical user interface* component manages the display and responds to the users requests either in the form of mouse click, button selection or natural language dialogue.
- The *Dialogue Agent* undertakes dialogue with the user and has the ability to understand natural language queries, and formulate a response. The role of the DA is to capture information in the form of attributes to answer questions about the specific domain. The DA obtains its own dialogue from a number of scripts, each representing a given context. Each script contains a series of rules and patterns which have been semi-automatically scripted by humans who have gained knowledge of the domain through the knowledge engineering process. An example rule for dealing with an employee who expresses confusion in natural language about the question being asked by the GO-CA is shown below:

```
<Rule_04>
p:50 *<confused-0>*
p:50 *<confusing-0>*
p:50 *<sure-neg-0>*
p:50 *<sure-neg-1>*
p:50 *help*
p:50 *not *<understand-0>*
r: How can I help you
```

where *p* is the strength of the pattern and *r* the response. Patterns can contain wildcard elements "*" which will match with one or more consecutive characters. In addition, the macro "<confused-0>" enables the scripter to incorporate stock patterns into a rule [9].

The construction of each GO-CA comprises of the following high level stages:

- Knowledge Engineering – Capturing and structuring of knowledge about the specific problem domain and the organisations culture and beliefs.
- Design and development of a rule based expert system to achieve user goals.
- User Interface Design (Web/ Intranet/ mobile / PDA/ voice recognition).
- Dialogue script design and organisation.
- Internal database integration for data validation or information extraction.
- Custom-built architecture involving the integration with the organisations existing systems.

4   Case Study 1: Student Debt Advisor

The first case study describes a GO-CA called Adam which was developed by Convagent Ltd [17]. Adam is a UK University Student Debt Advisor and is highly focused on providing advice to students by finding out why they have not paid their debts and offering advice on how they can find ways to pay them. The nature of support offered has required Adam to cope with upset and angry students, and deal with abusive language. Student debt is a growing concern in the UK with the average cost of a three-year University degree set to range from £38000 upwards [18]. 'Adam'

was designed to stimulate the behaviour of student debt advisors within the university and give clear and consistent advice on what help was available to students who had debt problems.

The GO-CA was designed to capture all the relevant information from the student's natural language dialogue in order to help them with their debt. Knowledge trees were used to represent the policies for student debt management within the organization. The GO-CA, guided by the knowledge tree had awareness of the type of attributes that could be provided by the student, such as database attribute names, and values, dates, and comparative terms. In order to achieve this, a different script was developed to identify the presence or absence of each possible relevant attribute and its value that would allow the GO-CA to guide the user through the University student debt advisor process. Example dialog between a student (Tom) and Adam can be seen in Figure 2.
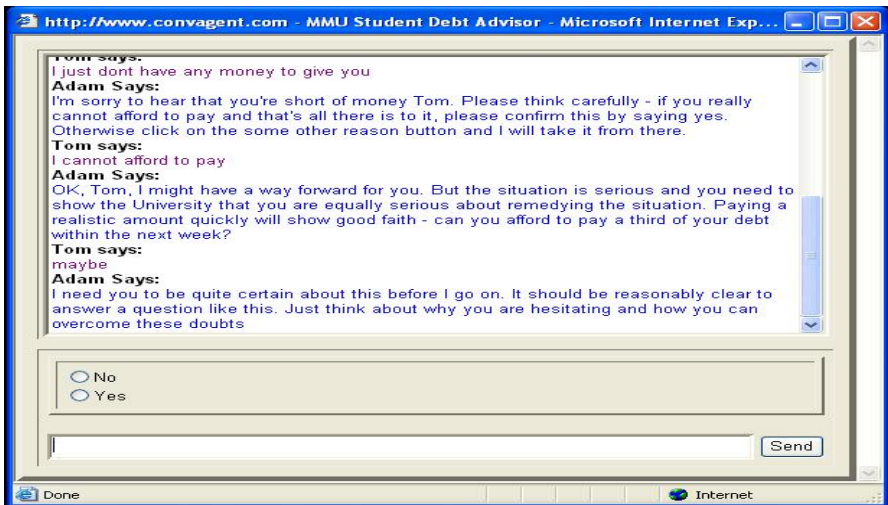

Fig. 2. Sample interaction with 'Adam'

4.2   Evaluation

The GO-CA, Adam was subjected to usability testing with 200 undergraduate students before going live. Students were given one of 5 scenarios which were designed from real life student circumstances of debt. For example, one such scenario is about a student called Tom who has got behind with his fee payments, has sent a late cheque and in the mean time has received a warning letter from the University. This scenario was selected as it is one of the two most common reasons for students to call the University human helpline.

Students were asked to provide feedback of their experience of 'Adam', using a questionnaire. 80% of students were happy with the speed in which Adam elicited the responses; 70% of students were satisfied by the advice provided by Adam and 75%

thought that Adam answered any additional queries that they had satisfactory. In the last part of the questionnaire, students were asked, "If you needed advice about your student debt problems, where would you go first?" 47% of students stated that they would use Adam in instead of visiting the University Finance Office while 20% stated they would use Adam instead of telephoning the University Finance Office. From this initial testing phase, the majority of student's comments were positive. For example "He needs to be a bit friendly, I felt like I was seeing a real Advisor…. No wait that's a good thing!" However, the testing phase identified a number of areas where Adam could be improved such as, a better understanding of student life experiences and the use of slang and mobile text talk. These ideas were then incorporated into the 'live' version of Adam which is currently being run at the University. Further studies are on-going and feedback obtained from the GO-CA log files enables 'Adam' to continue to learn and relate to life as a student.

5   Case Study 2: HR Bullying and Harassment Advisor

The second case study will describe how a GO-CA was used to act as an advisor for employees on 'Bullying and Harassment in the Workplace' policy and procedures in a large organisation [19]. In the UK, no single piece of legislation addresses the problem of bullying and harassment [20]. Rather, organisations are required to implement a number of different laws which protect employees from harassment due to various causes. This can result in the development of complex policy documents. An individual may not be able to apply the policy or procedure to their personal situation. This type of policy will often require additional training and guidance to support for members of the organisation wishing to understand the reporting problem. The high support cost and fairly static nature of such policies means that they are suitable for automation using a CA. This would allow anonymous, consistent and appropriate advice to be available 24/7. Knowledge engineering was used to elicit the main questions in relation to bullying and harassment asked by employees within the organisation [19]. Then a rule base was then used to structure the bullying and harassment domain. Figure 3 shows a portion of the rule base for structuring knowledge about the 'I am being victimized' option. Through natural language dialogue with the user, the GO-CA looks for each possible relevant attribute and any associated value in the user's input in order to determine what rule in the rule base to fire and hence what advice to give.

Figure 4 shows an example of a dialogue interaction between an employee and the bullying and harassment advisor 'Adam'. The dynamic interface presents the user with three methods of communication: graphical or through natural language dialogue or in a mixed mode. The interface was designed so that more experienced users or those with disabilities were able to select options by simply clicking on their choices. Alternatively, natural language dialogue can also be entered in order to discuss a person's individual circumstances, explain concepts and offer advice at any stage in the process.
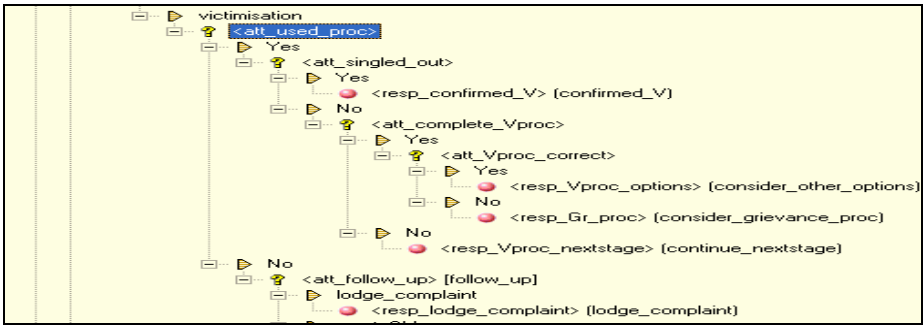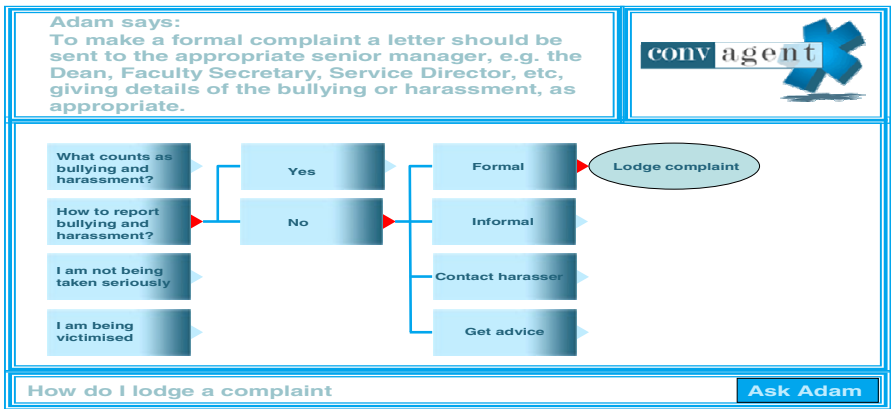
Fig. 3. Portion of rule base

Fig. 4. Bullying and Harassment Advisor

5.2   Evaluation

A representative group of 30 employees' were selected to evaluate the GO-CA. The group of 30 employees worked in different roles and for diverse companies, aged between 25 and 45. The group included both male and female members from administrative, academic and managerial backgrounds. A scenario of possible sexual harassment was developed along with a questionnaire to record feedback. The group were asked to read the scenario and then complete two tasks within the system. Users then completed an electronic evaluation questionnaire anonymously. Users felt that the system was intuitive and easy to use, giving high scores to ease of navigation and the ability to find information. On the whole, scores were high and the group found the system understandable and intuitive to use. 94% of users indicated that they had found the advice they sought without difficulty and one user commented that they "*did not need to 'learn' to use the advisor*". This is a very important benefit as users who seek to use this type of system do so because they want the same level of advice as provided by a human, without being constrained due to their technical ability.

6   Case Study 3 Conversational Intelligent Tutoring System

Case study 3, describes a Conversational Intelligent Tutoring System (CITS) known as Oscar which can dynamically predict and adapt to a learner's learning style during the tutoring session [21]. The system's goal is for a student to complete a tutorial which dynamically adapts to their learning style. Oscar CITS imitates a human tutor by leading a tutorial in natural language, intelligently analysing solutions and offering problem solving support rather than simply presenting the answers. In providing the learner with the most appropriate learning material for their learning style, Oscar CITS aims to improve the effectiveness of the learning experience and provoke a deeper understanding of the topic, and thus improve confidence. Learning styles are central to the Oscar CITS, so development started by considering the *Index of Learning Styles* (ILS) model [22] which was developed to describe the learning styles in engineering education and suggests different pedagogical styles to address learners' needs. Following an initial study, a subset of the best ILS predictor questions for each learning style dimension was then considered during the development of the Oscar CITS. The domain of SQL was selected due to its wide use in undergraduate courses. Tutoring revision scenarios were designed based around the syllabus and the database lecturers' experience of revision tutorials. Each revision question was mapped to the ILS model using the model's descriptions of indicative behaviour, such as a preference for theoretical questions. Following on from the tutorial design, the dialogue for the tutorial was scripted using Convagent's GO-CA [17]. Overall, there were 38 contexts containing around 400 rules which demonstrates the complexity of developing a CITS. The full architecture and methodology for developing Oscar CITS is described in [21]. Throughout tutoring, the CITS logs information including the behaviour of the participant, their knowledge of the topic and the adaptation employed.

6.1   Evaluation of Oscar

An initial study comprising of 35 native English speakers was conducted to investigate the adaptation of tutoring to an individual's learning styles for three ILS dimensions. Each participant had had previous experience of an undergraduate SQL course. Initially, participants were asked to complete the formal ILS questionnaire followed by a multiple choice test to assess existing SQL knowledge. Next, they engaged in a personalised tutoring conversation led by Oscar. During tutoring, each participant answered questions and completed various tasks in SQL. Depending on their level of knowledge, participants were exposed to various resources and given hints and help as required. Finally, each participant was asked to complete the same multiple choice test to measure their learning over the session. The results of the gave an accuracy of predicting learning style on the Sequential/Global dimension of 75-80%. Overall, the results have shown that the Oscar CITS tutoring seemed to help learning as all participants who did not initially achieve full marks improved their test scores by an average of 25%.

7   Conclusions and Future Research Directions

User evaluation of each of the 3 diverse GO-CA's has been positive and has illustrated that such an agent can help a user achieve a sophisticated goal through natural language dialogue. Whilst the three cases described in this paper highlight the clear benefits of using GO-CA, there are a number of research opportunities to improve the methodology. Firstly, capturing the domain knowledge and then scripting a GO-CA is a labour-intensive process. Changes in the environment and/or in the knowledge would effectively mean that new scripts would need to be created meaning the maintenance of such systems is high and the process of incorporating new scripts could lead to conflicting rules in the rule base. An alternative approach proposed by O'Shea et al [23] replaces the pattern matching approach to scripting by instead making use of sentence similarity measures. Sentence Similarity Based on Semantic Nets and Corpus Statistics [24...26] is a measure that focuses directly on computing the similarity between very short texts of sentence length. In the proposed methodology, sentence similarity measures [24...26], are employed to examine semantics rather than pattern matching. Initial results have indicated a substantial reduction in the GO-CA's rule base as scripted patterns in all rules are instead replaced with one or two natural language sentences. Secondly, evaluating any form of text based conversational agents is difficult to achieve as there is no clear scientific benchmark. Early evaluation methodologies such as PARADISE [27] were designed for speech-based conversational agents where the overall goal in the conversation was user satisfaction. A more scientific approach has been developed by O'Shea et al [29] who has developed a benchmark data set of 65 sentence pairs with human-derived similarity ratings. This data set will be a viable measure in the evaluation of *all* text based CA's in the future.

References

1. Massaro, D.W., Cohen, M.M., Beskow, J., Cole, R.A.: Developing and evaluating conversational agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.) Embodied Conversational Agents.MIT Press, Cambridge (2000)
2. O'Shea, J., Bandar, Z., Crockett, K.: Systems Engineering and Conversational Agents. In: Intelligent-Based Systems Engineering. ISRL, vol. 10, pp. 210–232. Springer, Heidelberg (2011)
3. Crockett, K., Bandar, Z., O'Shea, J., Mclean, D.: Goal Orientated Conversational Agents - the rocky road to commercialization. In: IEEE WCCI, Barcelona, pp. 1820–1827 (2010)
4. Mauldin, M.: Chatterbots, Tinymuds, And The Turing Test. AAAI, Menlo Park (1994)
5. Turing, A.M.: Computing Machinery and Intelligence. Mind 59(236), 433–460 (1950)
6. Weizenbaum, J.: ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine. Communications of the Ass' for Computing Machinery 9, 36–45 (1966)
7. Colby, K.: Artificial Paranoia: A Computer Simulation of Paranoid Process. Pergamon Press, Oxford (1975)
8. A.L.I.C.E. AI Foundation, Inc. 2011, The A.L.I.C.E. Artificial Intelligence Foundation (2011), http://alice.pandorabots.com/ (viewed January 13, 2011)
9. Michie, D., Sammut, C.: Infochat Scripter's Manual. Convagent Ltd., Manchester (2001)

10. Cassell, J., Sullivan, J., Prevost, S., Churchill, E.: Embodied Conversational Agents. MIT press, Cambridge (2000)
11. Cassell, J., Gill, A., Tepper, P.: Coordination in Conversation and Rapport. In: Proceedings of the Workshop on Embodied Natural Language, Association for Computational Linguistics (2007)
12. Massaro, D.: A framework for evaluating multimodal integration by humans and a role for embodied conversational agents, pp. 24–31 (2004)
13. Owda, M., Crockett, K., Bandar, Z.: Conversation-Based Natural Language Interface to Relational Database. In: IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 363–367 (2007)
14. Pudner, K., Crockett, K., Bandar, Z.: An Intelligent Conversational Agent Approach to Extracting Queries from Natural Language, World Congress on Engineering. In: International Conference of Data Mining and Knowledge Engineering, pp. 305–310 (2007)
15. Cha, H.J., Kim, Y.S., Park, S.-H., Yoon, T.-b., Jung, Y.M., Lee, J.-H.: Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 513–524. Springer, Heidelberg (2006)
16. VirtuOz (2011), http://www.virtuoz.com (accessed January 13, 2011)
17. Convagent Ltd. (2011), http://www.convagent.com (accessed January 13, 2011)
18. University funding (2011), http://www.bbc.co.uk/news/education-11483638 (accessed January 12, 2011)
19. Latham, A., Crockett, K., Bandar, Z.: A Conversational Expert System Supporting Bullying And Harassment Policies. In: Proceedings of the 2nd International Conference on Agents and Artificial Intelligence, pp. 163–168 (2010)
20. Bullying in the workplace (2011), http://www.direct.gov.uk/en/Employment/Employees/DiscriminationAtWork/DG_10026670 (accessed January 13, 2011)
21. Latham, A., Crockett, K., McLean, D., Edmonds, B.: Predicting Learning Styles in a Conversational Intelligent Tutoring System. In: Luo, X., Spaniol, M., Wang, L., Li, Q., Nejdl, W., Zhang, W. (eds.) ICWL 2010. LNCS, vol. 6483, pp. 131–140. Springer, Heidelberg (2010)
22. Felder, R., Silverman, L.K.: Learning and Teaching Styles in Engineering Education. J. Engineering Education 78(7), 674–681 (1988)
23. O'Shea, K., Bandar, Z., Crockett, K.: Towards a New Generation of Conversational Agents Based on Sentence Similarity, Lecture Notes in Engineering, In: Advances in Electrical Engineering and Computational Science, vol. 39, pp. 1100–1876. Springer, Heidelberg (2009)
24. Li, Y., Bandar, Z., Mclean, D.: Measuring semantic similarity between words using lexical knowledge and neural networks. In: Yin, H., Allinson, N.M., Freeman, R., Keane, J.A., Hubbard, S. (eds.) IDEAL 2002. LNCS, vol. 2412, pp. 111–116. Springer, Heidelberg (2002)
25. Li, Y., Bandar, Z., O'Shea, J., Mclean, D., Crockett, K.: Sentence Similarity Based on Semantic Nets and Corpus Statistics. IEEE Transactions on Knowledge and Data Engineering 18(8), 1138–1150 (2006)
26. Li, Y., Bandar, Z., McLean, D.: An approach for Measuring Semantic Similarity Between Words Using Multiple Information Sources. IEEE Transactions on Knowledge and Data Engineering 15(4), 871–881 (2003)
27. Sanders, G.A., Scholtz, J.: Measurement and Evaluation of Embodied Conversational Agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.) Embodied Conversational Agents (2000)
28. O'Shea, J., Bandar, Z., Crockett, K., Mclean, D.: Benchmarking short text semantic similarity. Int. J. Intelligent Information and Database Systems 4(2) (2010)