Natural Language Generation (NLG) in Artificial Intelligence Systems

Addanki Adithya
Department of Computer Science, The University of Akron.
Akron, Ohio-44304.
aa207@zips.uakron.edu

Abstract

This paper discusses how Natural Language Generation emerged as a field of Human Computer Interaction in Artificial Intelligence Systems, its origin, building blocks and basics of architecture. The paper further dwells into the problem at hand; making AI systems interaction effective with a personal style, trying to bridge the gap between humans and computers in a conversational manner. We look how the efficiency of few systems under study increased with the integration of NLG to them, what are the issues with current innovations, what could be done to promote AI to the level where it is difficult to differentiate if the discourse is generated by a person or a machine.

Keywords: NLG- Natural Language Generation, HCI- Human Computer Interaction, NLU- Natural Language Understanding, DM- Dialogue Management, ASR- Automatic Speech Recognition, TSS- Text to Speech Synthesis

1. Introduction

Intelligent Assistant systems like Apple's Siri, Microsoft's Cortana, and Nuance's Nina with their significant others have now made the things simpler [3]. A computer system would be considered as a sophisticated machine when it can communicate its results to the user successfully. An Intelligent system is used by a wide range of users from naïve to expert, but the sophistication of the system is of no use if the user fails to understand what the system is capable of and how to get the desired task done [4]. This gap leads to infrequent

use of the system due to lack of proper knowledge on the formal language used to communicate with the system. Hence NLP took its emergence into the field of AI making this communication as informal as possible [Figure 1.1], thus allowing naïve users get the best out of these intelligent systems with knowledge that is barely in substitution for computer languages.

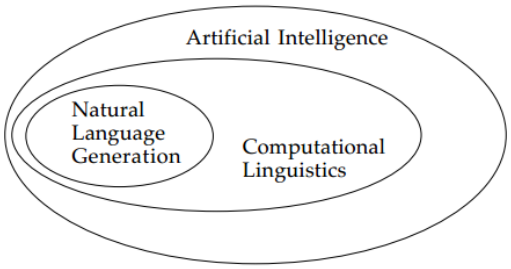

Figure 1.1: NLG's association with AI [12]

NLG is the area that motivates research in developing methods for making a computer respond to user through natural languages; human languages. NLG is found to be making a system more interactive compared with other categories of HCI. As sophistication of Interpretation systems increases; the analysis of reasoning behind selection of a choice might help determine the goals of speaker [4].

2. Problem

Three classes of decisions have to be carefully answered when it comes to making the system sophisticated as the user expects; what to communicate (maintain brevity and not re-stating the obvious), when to say what, and which words and syntactic structures best express its intent [4]. Research on interpretation describes limitations on options available whereas research in generation specifies which option is better and why, under those circumstances/situations/scenarios [1].

This paper aims to study the features of computational linguistics associated with natural language generation in AI systems (Speech and Text); the systems that have been implemented till date and the problems they face and what has been done so as to overcome these issues. When a connected text is to be generated the issues of discourse coherency and discourse structure are important. There is also a need for organizational framework that provides appropriate definitions. Let us now dwell into these concepts further and look at the design and operational architectures where applicable.

3. Background

Natural Language Generation could be viewed as the task which generates human languages using a knowledge base. A knowledge base is a typical database consisting of axiomatic and inference rules which help deduce sentence structures that are semantically valid. The primary objective of a Language Generator is to know how to put a concept into a word form that is understandable by the user. Most of the interactive systems create dynamic sentences based on interaction with the user to meet a communicative goal [1]. In order to generate a text or speech that is natural and non-repetitive there should be a planning and organization goal in place. Given below are the building blocks of NLG systems.

Content Determination: This module has to determine what should be the content, what information is to be included in the discourse. Once content has been finalized it undergoes structural and syntactic analysis for validation by succeeding modules.

Structure/Discourse Strategy: Organizing the information to be conveyed could be a difficult task depending on the scenario the interaction is dealing with, ranging from simple horoscope system where text is taken from a template to an advanced system that interacts with user and provides a dynamic service such as creating an appointment with a doctor.

Aggregation/Discourse Coherency: Several sentences in a discourse might sometimes lead to a flow that is discontinued and might not be of any interest to the user. Such circumstances demand for a more intelligent system that is capable of merging the content so as to fuse the sentences and create a different form of discourse that conveys the same information.

Lexical Choice: Ability to determine the choice of words in establishing the strength or significance of the concept being delivered as text or speech.

Referring Expression Generation: Identifying the objects or regions of interest as a part of something whole, and concepts as things having a meaning in a context [12].

Surface Realization: The concluding process which generates the resulting output; the actual result should be valid and well-formed according to rules of syntax, semantics, morphology and orthography.

A non-trivial language generation system requires the following [Figure 3.1]:

- A non-linguistic reasoning component
- A representation of the discourse
- A notion of a model and
- A representation of available linguistic resources; both grammar and lexicon encoded.

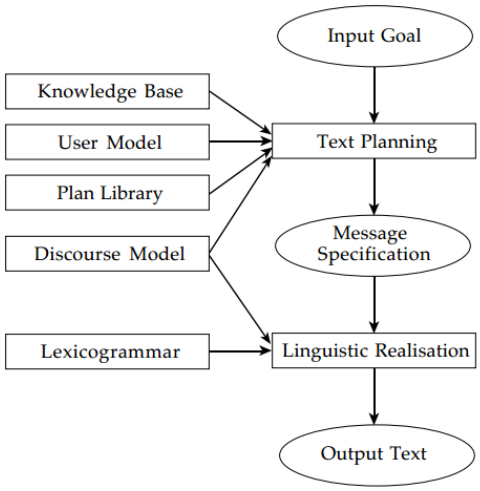

Figure 3.1: Standard Architecture of NLG systems [12]

Depending on the scenario and domain the intelligent system is supposed to be a solution for, there might be additional operational modules that assist language generation in those systems. The following decisions have to be made by any NLG system to be considered complete.

- The text content and what information should be omitted?
- Organization of content into a coherent discourse
- Tone and degree of formality
- Decomposition into sentences
- Choice of syntactic structures
- How entities should be described and Choice of words

A typical Dialog Management System has a Reasoning Module with a Knowledge Base and an associated rule base to form the basis of axioms, reasoning algorithm, and uniform interfaces to dialogue management module & third party applications.

3.1. Hard Problem- Complete AI

Interaction systems have to deal with many problems such as understanding the user's speech, occasionally a translation of user's input to another language, co-reference resolution; understanding which words refer to which objects, convert information into phrases of text/speech, identifying parts of speech, reasoning, question- answering, etc., These problems require just as much capability as human brain to process and communicate in a natural way, hence Natural Language Generation is termed as Complete AI problem.

Though natural language generation started its life in 1980s (McDonalds Mumble, HAM RPM, Appelt's Telegram and Kukich's stock market report generator), the first and foremost system that used natural language generation in real life scenario and gained reputation is FoG, a weather reporting system in early 90's that generates weather forecasts from weather data. Following the success of FoG, this approach of interaction through NLG has found its extensions into financial and business sectors to summarize the data and analyze huge amounts of sales data. Several major pieces of research works, emergence of NLG and the ubiquitous and increasingly sophisticated programs on which the systems are built on, directed the interest of research and development towards NLG over the last two decades. The next sections of the paper give a brief study of a few interactive systems and conclude with what is needed to improve the interaction to make it seem more naturally communicative than they are at present.

4. A Study on Interactive Systems

4.1.Personal Digital Assistants

A personal assistant has to adapt to the distinct demands of its master and increasingly pay attention to the user's preferences and routines. The common features expected from a Digital Assistant are flexibility, simplicity and ease of use [4]. These requirements are best approached with voice based interaction, which doesn't require much cognitive efforts, attention or memory necessities on the part of the user who interacts. The scenarios where

these systems are used range from simple command to surf the internet, to setting up a meeting or managing the calendar. A feeling of natural and realistic interaction is required to make the communication more effective and efficient.

Technology Integration in the last few decades has been the focus of extensive academic research. Even with the robust technology we have today, the context and personal behavior of PDA is still at the start. Though not really personal Apple Siri, has given a personal touch to AI, through its sassy tone. It tries to follow the user's orders, names you, it is just not a personal assistant; it has a character. Natural language understanding and dialogue management of the technology pipeline require the biggest amount of augmentation. Despite the promise of being context-aware Intelligent PDA in such pure sense are still missing [4]. Some areas of improvement are given below.

Extended Dialog History: Involves keeping a dialogue history to keep track of conversations and adapt to users preferences over time.

Improved Context Awareness: Using sensors to make the PDAs aware of the environment we are in to make it know more, improve systems reasoning skills and provide augmented HCI.

Dynamic System Adaptation: Proposal involving use of multi-agent architecture so that the output from one agent could be used as input to another thus resulting in more dynamic system that adjusts the dialogue accordingly.

Supported Task Hierarchy Design: Information state paradigm and stochastic processes are not found to be useful as they need training data in order to start with building the system. Task-based paradigm is proposed as the best suited technique, depending on size of dialogue, initiative and conditional execution steps, these task hierarchies can be quite complex.

4.2.NLG for Documenting Simulation Results

Simulation experiments generate large datasets which are analyzed to make decisions. NLG in this approach is used to check the experiment data, generate highlights and summarize the results, thus reducing the cost and complexity of the studies [9]. Of the large datasets generated through simulation, only a small subset of results are interesting to the

user, reporting these results could be time consuming. Adding a modeler that helps adding interpretation and refinement to the generated text, we could have a solid initial report to start with that does not allow the user to digress from what is intended to be included in the report. These reports have shorter body that gives the summary of conclusions and larger appendices that provide detailed information about the model and the experiment. Most of the naïve users need shorter yet concise and brief versions of summary included without excessive statistical analysis. NLG systems in this category have their primary goal as identifying a clear communication goal. Inputs to the program would be different internal and external factors in which the system resides under different conditions like number of machines, amount of inventory, sales, revenue, arrival rates etc [9]. NLG module converts the summary into a report adding the extra meta-data regarding these said factors. The following are typically different steps in documenting simulation results through NLG.

Data Analysis: Analyses the feasible and infeasible configurations for each situation and sends these to content determination for identifying information to be included in the report.

Content Determination: Simple rules that specify when to discuss the feasible and infeasible solutions, selecting interesting information for the user. Content Determination can be viewed as optimization problem where we have to maximize the information that is to be conveyed within the given page constraints.

Micro Planning and Aggregation: Combining multiple sentences into one to avoid repetition, choice of words has an impact in this area.

Realization: Applying grammar rules to combine words including word order to form sentences correctly and meaningfully.

4.3.Personal Narrative for children with Complex Communication Needs

A system that is used by children with complex communication needs at school, to create day-to-day routine stories. NLG is used to create a draft version of the child's daily activities and the child is allowed to edit it to suit his needs [5]. These tools are targeted at children with disabilities, who could not communicate better because of cognitive issues. Most of the existing systems deal with small practical and real life goals like uttering "I am

thirsty", like a passive communicator at the press of a button [5]. Personal narratives about his or her life on a daily basis is a kind of conversational narrative, requiring specific tools to help challenged people to participate in sharing their experiences; a story if the user gives an emotional connection to it. These devices have cognitive demands and require physical stamina on the user's part, because of the interactive mechanism that requires a sequence of operations to produce the required output. [Figure 4.3.1] displays a system HWST which has RFID reader that is used for voice message recording, a visual interface to display the possible choices for output which can be uttered by the system, access switch in the user's head to choose the options, long range RFID for identifying the environment conditions. RFID tags are used to track location so as to use geographic information in the story, like where the incident happened. Teachers and parents could also record their messages, any divergence in the routines is identified by comparison against the student's timetable. At the end of the day, the system selects 5 top most interesting events and provides the user an interface to edit and store them into memory. Once the editing is done, the NLG system translates the story into texts describing those events of interest. The challenging task is to choose connectives, time phrases, lexical choices and pronouns based on dialogue content. In case a child wants to reorder the events based on his or her interest levels, the NLG system is to be provided with a more advanced interface to do the same. A few major impediments are delays caused by starting the system, limited battery life, difficultly in tracking the teachers or parents with whom the conversation took place(as it required swiping magnetic cards through the interface or tagging/ annotating their speech with their names). Location tracking issue was later solved using WiFi enabled GPS system that continuously signaled the users geographic co-ordinates. Restricting motor skills make few tasks such as entering a few words tedious, requiring cognitive efforts and time consuming, the NLG system is expected to reduce such difficulties and give the child as much control

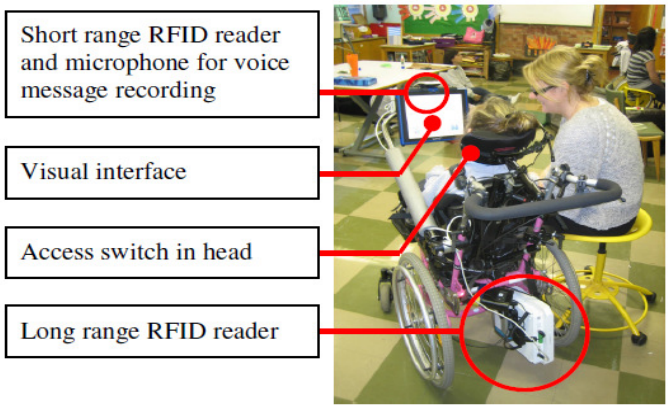

Figure 4.3.1: How was School Today: System for children with CCN[5]

as possible with minimum efforts from the user [5]. Once the stories are developed, they are tagged with date and time to give a sense of the child's history when required, the child can refer to these later and lengthen the story. It is also expected to include a provision to give the ability to support complex conversations like, "guess what happened at school today?" One more interesting challenge is that the children are more inclined to interact with other children of their age, there arises a new issue to determine what story to say in order to attract the interest of the other people involved in the conversation. Security and privacy is also a concern in this modern world; the recorded events and messages should only be given access to the children or their parents, or else these might be misused if fallen into hands of malicious outsider.

4.4.Song Lyric Generator

Lyric generation is also one of the difficult problems to solve as it involves both creativity and inspiration. Music is one of the oldest art that has existed since caveman music to pop, rock and jazz cultures. Music can be quite uplifting and relaxing for humans, often it is used for soothing the nerves and stress out. Our world is becoming more of a global village as a result of growing technology thus bridging the gap between people around the globe [11]. To reduce the task burden in writing lyrics for song writers and musicians song lyric generators have been under research and development. The implemented systems generate the lyrics based on musical beat, melodies and rhymes identifying the syllables and the stress on each word. A few systems had databases for about 40,000 existing lyrics and new lyrics are interpolated using n-gram model approach. Once a couple of lyric verses are generated these are merged according to song structure and layout but the problem found was that the lyrics were not entirely meaningful and did not relate to the required theme.

WASP (Wishful Automatic Spanish poet) was a system that generated lyrics based on seeds; base words in the lyrics given as input by the user, through forward reasoning [11]. These were not much successful and appealing either. A research conducted at the University of Mauritius, included 300 songs from three different genre (love, pain and cause) with 100 songs each, studied the structures of songs, and analyzed the chorus,

verses, bridges and hooks. Then it was known that it is not just creativity and inspiration which are required for lyric generation, and then identified the missing process as assembly. The organizational structures are studied stored as n-gram sequences, Parts-of-Speech rules were established to tag the words in lyrics. The user starts by selecting the theme of the song, then giving a few input seeds to the lyric generator and the system asks for a catchy verse that has to be used for repetition, finally the user has to select the structure of songs based on chorus and verse. The lyrics generated were evaluated using human without their knowledge, and 52% of them classified the songs as existing i.e., natural lyric. These limitations might be due to the natural ambiguities (classification of words in a verse) in natural languages which are to be eliminated using more intelligent systems. These might possibly and partly be eliminated using word sense disambiguation [11].

4.5.Reasoning Systems

These systems like question answering systems return a natural language generated output, but it is associated with a reasoning as how and why the result is deduced. It explains the user how the system reached a point of conclusion about the result. Expert system users are not programmers for majority, they are a group of users who understand natural language and are less familiar with computer languages. The primary objective of the system is to offer the users a consultation that confirms their own observations and diagnosis in the domain (can be viewed as verification with a standard that is accepted). Depending on the result of an expert system the user might choose to accept or deny his observations based on the reasoning the system has followed to prove a point. The inference process might often require thousands of rules to be triggered, that in a few cases it might be very valuable to retrieve a textual explanation to find out the trace where the error has occurred.

4.6.Question Answering (QA) Systems

An issue that was mistaken to be a primary parsing problem, a user's question is translated into a representation that is understood by the system, the underlying knowledge

base is searched for relevant answer and is represented to the user in tabular format. Many users who are primarily naïve need to familiarize the database as a whole before interacting with the QA system; what information is available in the database, what format is it in, what specific terms mean in a context ex. Market price and selling price.

Questions in spatial domain require more than just a search through knowledge base, the questions which require a description of an object may be answered as brilliantly or succinctly as possible. These questions are called meta-level questions and these are categorized into classes; requests for definitions, available information, differences between objects, object descriptions. A language generation component is required in this picture to determine what to say, the correct level of detail to be provided and the organization of the detail; this varies from user to user among possible different domains.

5. Current Issues with NLG in AI Systems

Given a discourse goal, an NLG system should provide an organizational framework to generate text/speech that aims to meet the required goal. Language generation has to formulate reasons regarding why that particular option has been chosen, out of possible many other options, to provide the best utterance for a given situation. Choices that the NLG systems face include options in content, textual shape & structure, and transformation of the message. Grammar rules are used to make decisions about both lexical and syntactic choices, and to reach a syntactic structure that best expresses its content. A few prevalent issues with NLG systems are discussed in this section.

Word-Sense Disambiguation: It is one of the most important preprocessing steps of Natural language generation whose accuracy depends on the machine readable dictionary and annotated corpus [8]. The context where the word appears decides the meaning of the word in that sentence, for example, the word “bank” has a noun and verb meaning, depending on the context where it appears, and alongside with water or an airplane, it takes different meanings. A graphically represented network called as Semantic network is used in order to help disambiguate the sense of words in a sentence. Different approaches have been proposed, one of them is gloss overlap that starts with word pairs of content words and tries to find the overlapping word structures that had the highest frequency, which

indirectly means that the two words we are trying to pair up are inherently coherent in nature. Emphasis would be on getting better connectivity that helps in making better judgment and decision making.

Executing Commands: It is said that "it is a long road from taking commands to its fulfillment", Dictation and command recognition demand to hunt for parallel methods which balance vocabulary, accent and context needs. Inputs and responses could vary in variety of ways no matter how restricted the domain is. For instance, when an agent is expecting the user to respond to it with a "yes" or "no", possible interactions could be "yeah", "yes that is correct", or "yep". When such simple interaction needs an extensive long list of phrases, it is hard to imagine how big the knowledge base should be to have a personal and intelligent assistant [3].

Context Awareness: Software agents have taken context awareness to a level that has never been attempted before, for instance Siri, a system that has a tone, a character; it tries to know more about you than you know about yourself, patterns, tastes, choices and preferences. This enables the systems to be equal partners as humans in the interaction, finding the right time for system's proactive involvement [3] [4].

Dialogue History: Even with such agents with great context awareness there are some places where it needs further polishing, such as when in a conversation, referring to objects and resolving the references to subjects or objects in the conversation. Keeping track of the conversation and maintaining a history helps the system to understand its user better [4].

Emotion Recognition: Emotion awareness is required to maintain an adaptive dialogue strategy. The prosodic features of the speech signal are extracted, usually pitch, formants, intensity, jitter, harmonicity, minimum and maximum values along with deviations. Emotions such as anger, fear, happiness and sadness features are identified, it was observed that largest contrast is between anger & happiness and neutral state. Also male and female voices exert different pitches, they also express their emotions with different intensities. Recognizing the emotion of the user is must have feature in AIs such as personal digital assistants [4].

Accurate Knowledge Base: A knowledge base should be complementing the scenario for which the NLG system is targeted at. Having a vast knowledge base that is irrelevant to the discourse task at hand is of no use. A dynamic knowledge base might also be helpful in cases where the discourse strategy requires frequent changes based on inputs from multiple agents i.e. sensors. Identifying the knowledge base that is sufficient to our discourse task is difficult, it should be domain specific and appropriate for reasoning with a level of detail that offers the required sophistication to the user [2]. Users should also be able to add and remove information from the knowledge base. An interface should be provided that inquires and reasons, achieves the result where the text generation module takes over.

Information Representation: Information has to be presented to the user in an effective and conceivable manner. It is a bidirectional requirement which expects proper information from the user's end too. When a user asks a PDA to "set up a meeting with Lisa at 9:00 pm on Saturday", the text is synthesized, parsed and then the AI system should be able to determine that Lisa is a person who is one of the user's contacts, 9:00 pm is a time slot in the evening and Saturday is the weekday on which the user wants to meet with Lisa. Once the system identifies this information, it has to interact with the calendar application in the backend and then creates a graphic user interface that pops up displaying the meeting schedule on Saturday. Note that several internal processes were hidden in this example such as checking if the user has any other appointments at that time, or resolving which Lisa is the person referring to in case he has multiple entries in the address book. Though information representation is clear in such systems, these might require some innovation to provide for the physically or mentally challenged or people with Complex Communication Needs (CCN).

Measure of Success: The ultimate goal of any AI system is to deliver quality and useful response to its user within acceptable performance constraints. Satisfaction of the user is another measure of success, because it is just not enough to get the information required by the user sometimes, it might be the case that the information should be delivered in an effective way. Defining standard benchmarks for evaluation of various AI systems has been difficult to establish [1]. The challenge of meeting sophisticated real world dialogue

requirements combined with the complexity of specialized contextual instances denote that it would be difficult to design any standard for effective interaction and efficiency in NLG.

6. Future works and Conclusion

The main purpose of any AI system is to be more usable and give a user friendly interface, NLG was observed to be one of the top notch interaction mechanisms that seems to have a future. An ability to integrate various sources and negotiate solutions between sources and its user is a feature that is expected. AI has the capability to paradoxically expand the power by collapsing infinite possibilities into one single action [10]. The future of NLG lies in imbibing the natural characteristics of humans into AI with a truly autonomous agent [6]. However, technological developments and impediments are quite common in this area of research for now as the field is relatively novel. Natural and spoken dialogue systems are expected to become important assets of the future world. The development of these systems is highly dependent on the coherence of integration between technologies, techniques and interfaces chosen [7]. NLG is a discipline that involves problems dealing with choice, based on factors identified, requiring researchers to dwell into this area to make it naturally communicative.

7. References

[1] . McKeown, K. R. (1986). Language generation: Applications, issues, and approaches. Proceedings of the IEEE, 74(7), 961-968. doi:10.1109/PROC.1986.13575

[2] . Wang Weiwei, Lin Biqin, Chen Fang, & Yuan Baozong. (1996). A natural language generation system based on dynamic knowledge base. Paper presented at the Signal Processing, 1996., 3rd International Conference on, , 1 765-768 vol.1. doi:10.1109/ICSIGP.1996.567375

[3] . Tom Geller. 2012. Talking to machines. Commun. ACM 55, 4 (April 2012), 14-16. DOI=10.1145/2133806.2133812 http://doi.acm.org/10.1145/2133806.2133812

[4] . Milhorat, P., Schlogl, S., Chollet, G., Boudy, J., Esposito, A., & Pelosi, G. (2014). Building the next generation of personal digital assistants. Paper presented at the

Advanced Technologies for Signal and Image Processing (ATSIP), 2014 1st International Conference on, 458-463. doi:10.1109/ATSIP.2014.6834655

[5] . Rolf Black, Joe Reddington, Ehud Reiter, Nava Tintarev, and Annalu Waller. (2010). Using NLG and sensors to support personal narrative for children with complex communication needs. In Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies (SLPAT '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 1-9.

[6] . Maja Pantic, Nicu Sebe, Jeffrey F. Cohn, and Thomas Huang. (2005). Affective multimodal human-computer interaction. In Proceedings of the 13th annual ACM international conference on Multimedia (MULTIMEDIA '05). ACM, New York, NY, USA, 669-676. DOI=10.1145/1101149.1101299 http://doi.acm.org/10.1145/1101149.1101299

[7] . Lars Ackermann and Bernhard Volz. (2013). model[NL]generation: natural language model extraction. In Proceedings of the 2013 ACM workshop on Domain-specific modeling (DSM '13). ACM, New York, NY, USA, 45-50. DOI=10.1145/2541928.2541937 http://doi.acm.org/10.1145/2541928.2541937

[8] . Charhate, S., Dani, A., Sugandhi, R., & Patil, V. (2012). Adding intelligence to non-corpus based word sense disambiguation. Paper presented at the Hybrid Intelligent Systems (HIS), 2012 12th International Conference on, 173-178. doi:10.1109/HIS.2012.6421329

[9] . Curry, J., Weihang Zhu, Craig, B., Turpin, L., Bokhari, M., & Mhasavekar, P. (2013). Using a natural language generation approach to document simulation results. Paper presented at the Simulation Conference (WSC), 2013 Winter, 2116-2126. doi:10.1109/WSC.2013.6721589

[10] . Leopold, H., Mendling, J., & Polyvyanyy, A. (2014). Supporting process model validation through natural language generation. Software Engineering, IEEE Transactions on, 40(8), 818-840. doi:10.1109/TSE.2014.2327044

[11] . Pudaruth, S., Amourdon, S., & Anseline, J. (2014). Automated generation of song lyrics using CFGs. Paper presented at the Contemporary Computing (IC3), 2014 Seventh International Conference on, 613-616. doi:10.1109/IC3.2014.6897243

[12] . Robert Dale,.An Introduction to Natural Language Generation. http://clt.mq.edu.au/~rdale/teaching/esslli/part01.pdf