Application of Natural Language Processing (NLP) Techniques in E-Governance

Siddhartha Ghosh[1]* and Sameen S Fatima[2]

ABSTRACT

E-Governance is the public sector's use of information and communication technologies with the aim of improving information and service delivery, encouraging citizen participation in the decision-making process and making government more accountable, transparent and effective. Effective and efficient e-Governments deploy Information and Communication Technology Systems to deliver services through multiple channels that are accessible, fast, secure, reliable, seamless, and coherent. To implement better G2G, G2B, G2E and G2C services we should not only utilize ICT, we have to be also serious about implementing Natural Language Processing (NLP) Techniques to reach up to the masses and make e-governance successful one. This paper shows the need of applying NLP technologies in the field of e-governance and also tries to focus on the issues, which can be resolved very easily with the help of these modern technologies. It also shows the advantages of applying NLP in e-governance.

Key words: e-Governance, natural language processing, NLP, ICT.

1. Introduction

Good governance is an act of balancing the interests that cut across community, space and time. The task is multi-dimensional and requires the employment of tools that can articulate the process through meaningful participation of the common man. Once we talk about the common man we should also think about how they communicate and how they understand

If the major aims of e-governance are:

- To strengthen and upgrade the present infrastructure of information technology in the ministry.
- To bring in transparency the functioning by placing the information, to the extent possible, in the public domain.
- To interlink the functioning of different offices to bring about synergy in the aspirations and enable effective and timely decision-making.
- To develop public grievances and complaint system and eliminate the need of personal visits to offices by the public.

---

[1] Department of Computer Science & Engineering, G. Narayanamma Institute of Technology & Sc., Hyderabad–08, India

* *Corresponding Author:* (E-mail: mr_siddharth@yahoo.com, Telephone: +91-9885143429)

[2] BITS, Pilani - Dubai Campus, UAE, on lien from Dept. of CSE, Osmania Universty, Hyderabad, Andhra Pradesh, India

- To make available forms and information on the Internet and the facility of filling the applications from remote rural areas-web Enabled application system.

Then the next major point should be:

- Application of modern technologies by the e-governance (i.e. NLP techniques)

According to Wikipedia Natural language processing is a sub-field of artificial intelligence and linguistics. It studies the problems of automated generation and understanding of natural human languages. Natural language generation systems convert information from computer databases into normal-sounding human language, and natural language understanding systems convert samples of human language into more formal representations that are easier for computer programs to manipulate.

Speech synthesis: although this may not at first sight appear very 'intelligent', the synthesis of natural-sounding speech is technically complex and almost certainly requires some 'understanding' of what is being spoken to ensure, for example, correct intonation.

Speech recognition: basically the reduction of continuous sound waves to discrete words.

Natural language understanding: here treated as moving from isolated words (either written or determined via speech recognition) to 'meaning'. This may involve complete model systems or 'front-ends', driving other programs by NL commands.

Natural language generation: generating appropriate NL responses to unpredictable inputs.
Machine translation (MT): translating one NL into another.

Major tasks in NLP includes

- Automatic summarization
- Foreign Language Reading Aid
- Foreign Language Writing Aid
- Information extraction
- Information retrieval
- Machine translation
- Named entity recognition
- Natural language generation
- Optical Character Recognition
- Question answering
- Speech recognition
- Spoken dialogue management
- Text simplification
- Text to speech
- Text-proofing

In the next couple of points we will try to focus on how some of the above mentioned techniques could be utilized to enhance e-governance services. Before that we will have a look in to the major e-governance projects for any country.

2. Major e-governance related activities

Country portal
Country portal is a one-stop shop to all Government Business in any country. There are various governments departments and ministries in the Centre and various agencies. The Country portal is a single link to all the departments and agencies and also the state/sub governments.

Services those are offered under this category:

- Forms (Downloading)
- Directory Services (For E-Mail, telephones, Addresses, Contact Personnel of all Government Offices)
- Government Schemes
- Links to all other Government sites
- Links to Education
- Health Services
- Survey Reports
- Parliamentary Discussions
- Trade related Queries
- Legal sub-portal
- Grievance Handling
- File Monitoring
- State Government Links
- Employment
- Who's who of Government
- News etc

State Portals

State portals are the single window Government entry points of various States. On lines of the National portal the State portals incorporates all most the same solutions mentioned above as well as link to central Govt. resources.

EDI

Electronic Data Processing can be implemented in every level of central and state governments.

e-Biz

The e-Biz project is a single window for all Government to Business services. The proposed E-BIZ can incorporate the following services:

- Information about laws, regulation, codes and procedures involved in running businesses in the country
- Filing of taxation, returns, claims and refunds
- The registration and start up of a Company/ Business/ SME or any form of Organization
- Issue of Licenses, permits & certification
- Renewal of Licenses, permits & certification
- To facilitate the FDI
- Filing of reports, compliance of statutory provisions
- Availability of facilities like telephone, water, electricity, transport, posts
- Window for all Procurement by the Government
- Interactions with the Stock Exchange, SEBI and other regulatory bodies
- Recruitment of skilled manpower
- Forms required for various services
- Feedback on various initiatives / Discussion forums thereof
- Source of funding (banking & FI's)
- Obtaining Land and building approvals
- Compliance of labor laws and interaction with trade unions
- Issues related to IPR
- Technology transfer from R& D institutions to industry
- Getting feedback on market trends (market research)
- To act as the interface to international agencies

e-Procurement

Each Government Department, Agency, State Government at time procures material. With the use of ICT in procurement process, the procurement overhead cost will go down. There had been examples in the west

wherein Governments have reduced their costs by over 50% by choosing E-Procurement.
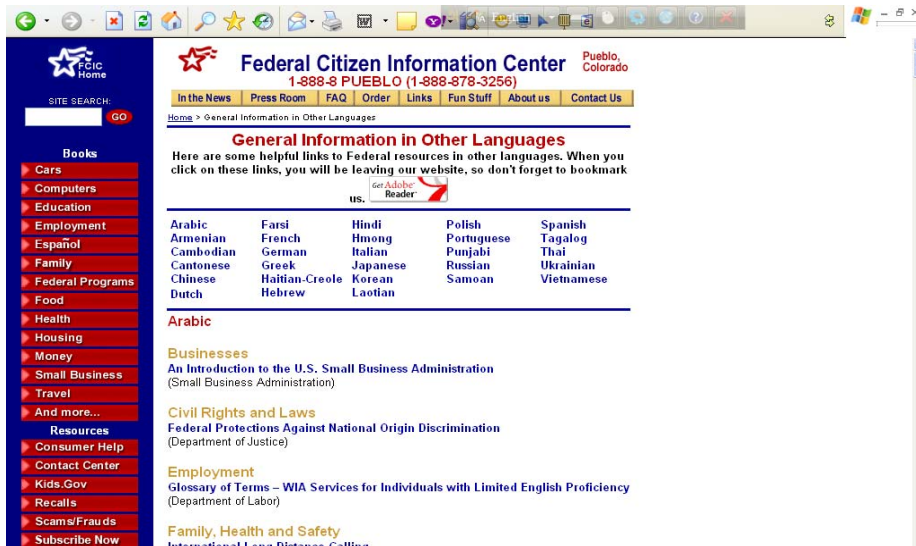

Figure 1: A multi lingual E-governance portal of USA govt.

3. NLP techniques in e-governance and how

NLP based e-governance systems are already in use some countries. As for example Natural Language Based Information Retrieval System Anveshak (The Quester) is a Natural Language based Information Retrieval system, which can efficiently and accurately provide explicit information in natural language text to the question intended to be queried on a certain document, developed by CDAC, India. But we fill that more initiative has to be taken in various level of state and central govt. to implement NLP based e-governance models or systems in every country. And mainly when it is matter of multi lingual countries like India then NLP may be the black horse of the e-governance.

As all of us know that e-governance is not only for educated people its for every one. Starting from a Govt. employee, up to a simple farmer in village any one can utilize the facilities offered by e-governance systems. Now lets us take an example – a farmer want to put the prices for his rice to sell through the kiosks available in village. If he is an educated person he can use the kiosk or internet connected computer, which is kept in the gram panchayet office, very easily. If he is not educated then a speech recognition (speech to text converter) system may be just work for him as the best friend.

As well as, if the speech recognition system is also enhanced with a speech translation system the even though the person may use his mother language like Gujrati and can quote the price in English. The NLP based machine translation system will be used here as a human friend. Blind people can use speech recognition based systems very easily. They can do form fill up for any govt. job application, death and birth registration, online application for including name in voter list etc. They can just utter the sentences; words and the system will help them to fed it in the required places.

Speech recognition (in many contexts also known as automatic speech recognition, computer speech recognition or erroneously as voice recognition) is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program. Speech recognition applications that have emerged over the last few years include voice dialing (e.g., "Call home"), call routing (e.g., "I

would like to make a collect call"), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g., a radiology report), domestic appliances control and content-based spoken audio search (e.g. find a podcast where particular words were spoken). Voice recognition or speaker recognition is a related process that attempts to identify the person speaking, as opposed to what is being said.

Machine translation, sometimes referred to by the acronym MT, is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another. At its basic level, MT performs simple substitution of words in one natural language for words in another. Using corpus techniques, more complex translations may be attempted, allowing for better handling of differences in linguistic typology, phrase recognition, and translation of idioms, as well as the isolation of anomalies.

Current machine translation software often allows for customization by domain or profession (such as weather reports) — improving output by limiting the scope of allowable substitutions. This technique is particularly effective in domains where formal or formulaic language is used. It follows then that machine translation of government and legal documents more readily produces usable output than conversation or less standardized text.

Improved output quality can also be achieved by human intervention: for example, some systems are able to translate more accurately if the user has unambiguously identified which words in the text are names. With the assistance of these techniques, MT has proven useful as a tool to assist human translators, and in some cases can even produce output that can be used "as is". However, current systems are unable to produce output of the same quality as a human translator, particularly where the text to be translated uses casual language.
Application

4. Use of natural language in govt. websites
As government websites are full of information, forms, data etc. it is the major vehicle of offering e-governance services. In this context the use of multilingual web portals, providing, e-governance services, can change the face of any govt. as well as it helps in overall development of any country or state. In India along with the central govt. so many state govt. offers multi lingual websites or local language based websites. In this regard we can name few of them which have already become very popular for their services.

Some good effort has been made or initiative has been taken form India Govt. side like the Hindi website of Ministry of Home Affairs (HA). Such effort is not found from smaller states level. Where as Govt. of Andhra Pradesh, Tamilnadu, Karnataka, and West Bengal have made good effort in this direction. A major reason may be the unavailability of the funds in this direction for all the states. But the Time has come when we should take up the issue of using local languages for developing websites for various purposes, like starting from basic information distribution to writing online exams in local languages.

5. Natural Language generation systems in e-governance
Natural Language Generation is a sub field of Computational Linguistics and language-oriented Artificial Intelligence research devoted to studying and simulating the production of written or spoken discourse. The study of human language generation is a multidisciplinary enterprise, requiring expertise in areas of linguistics, psychology, engineering and computer science. One of the central goals is to investigate how computer programs can be made to produce high-quality natural language text from computer-internal representations of information.

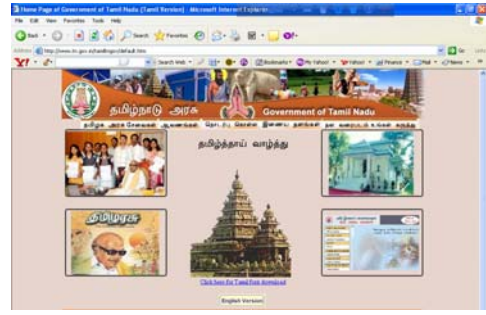**Figure 2:** Hindi website of Ministry of HA, India

Figure 2: Official Tamil Website of Govt. of Tamilnadu

Natural language generation often is characterized as a process that has to start from the communicative goals of the writer or speaker and needs to employ some sort of planning to progressively convert them into written or spoken words. In this view, the general aims of the language producer are refined into goals that are increasingly linguistic in nature, culminating in low-level goals to produce particular words. Usually, a modularization of the generation process is assumed which roughly distinguishes between a strategical (deciding what to say) and a tactical (deciding how to say it) part. This strategy-tactics distinction is partly mirrored by a distinction between text planning and sentence generation. Text planning is concerned with working out the large-scale structure of the text to be produced and may also comprise content selection. The result of this sub process is commonly taken to be a tree-like discourse structure, which has at each leaf an instruction to produce a single sentence. These instructions are then passed in turn to a sentence generator, whose task can be further subdivided into sentence planning, i.e. organizing the content of each sentence, and the final step of surface realization, i.e. converting sentence-sized chunks of representation into grammatically correct sentences.

Any e-governance system, which uses natural language generation techniques, can provide better service as a part of e-governance. As for example – a natural language generation based e-governance system can help the people to listen to the text as an audio files once any one selects them. Now the person may be educated but not familiar with that text but he can understand and speak it. Or blind for blind people it can read news, information, process, job vacancies information etc. For elderly people having short sight this tool will be a useful one. The NLP based question answer systems can just help the people like anything. A farmer who can not read text properly can even get help from a website, like how to use fertilizer in filed. How to grow new crop and how to water them using new techniques. Govt. institutions like India railway and Air India etc can utilize such systems for common people to make the services much people friendly.

The might and authority of the people of India, the largest democracy in the world, which pervades this Republic is represented by the President of the country. The official website of the first Citizen of India was recently revamped and transformed into a graphically appealing and completely dynamic website. Though the President's Office already had a website for the past few years, a need was felt to enhance the content and visual appeal as well as add some dynamic features which facilitate an interaction between the Hon'ble President and the common citizens. The re-designed Website, a joint effort of the President's Secretariat, the concerned NIC Cell and the Webservices Group at NIC HQ, is an online resource of information about the following: President of India, magnificence and glamour of Rashtrapati Bhavan, presidential retreats, ceremonial functions, Mughal Garden, Herbal Garden, a brief about the former Presidents of the nation, a collection of speeches / lectures / addresses / banquet speeches as well as press releases & recent events in multilingual format using dynamic fonts.

6. Concluding Remarks

Moving house, having a baby, setting up a business... there are numerous reasons why citizens or companies contact public administrations, and local government are under increasing pressure to improve the quality and speed of service delivery. More than offering just Web-based information, Time has come to give citizens personalized assistance and solutions to enquiries, entered entirely using natural language processing techniques. Its not a new concept but as the same time still now, we fill, these techniques are adopted by so many good e-governance providers. Time has come more R&D work has to be initiated in govt. level to take up this issue much seriously and making e-governance a common man's friend.

References

1. Sameer Sachdeva, (2002), *e-Governance Action Plan for India*, December'2002.
2. R.K.Khitoliya…et al; (2003), Potential application of IT based Techniques in Residential/Institutional Buildings, Seminar of Information Technology in Built Environment, 141.
3. Yogesh Baweja (2005), Does E-Governance means Good governance, *Yojana*, August 2005,Volume 49.
4. R. K. Khitoliya and Nitin Tanwar, E-Governance: An aid to the development of Isolated Rural Communities in Developing Countries, *Agriculture Today*, October 2005.
5. Anamika Raushina, (2005) E-governance for good Governance-The national E-governance Action Plan, *Kurukshetra*, Volume 53, No.5, March 2005.
6. Siddhartha Ghosh, Sameen S Fatima, (2007), use of local languages in Indian portals, *CSI Communication*, June'07 issue, pp- 4-12.
7. Siddhartha Ghosh, Sameen S Fatima, (2007), Retrieval of XML data to support NLP applications, ICAI'07- The 2007 International Conference on Artificial Intelligence Monte Carlo Resort, Las Vegas, Nevada, USA June 25-28, 2007.
8. Siddhartha Ghosh, Sameen S Fatima, (2007), A Web Based English to Bengali Text Converter, will be presented in The 3rd Indian International Conference on Artificial Intelligence (IICAI-07), Pune, India, December 17-19, 2007.
9. Standards and Architecture for E-Governance Applications - http://egovstandards.gov.in/localisation_wg/localisation_std
10. Unesco Portal - http://portal.unesco.org/ci/en/ev.php-
11. CDAC website - http://www.cdac.in/html/aai/anvesh.asp
12. Federal Citizen Information Center - http://www.pueblo.gsa.gov/multilanguage/multilang.htm?urlnet99
13. *Journal of E-government* - http://www.egovjournal.com/journal_archive_voli_issue1.htm
14. An E-Governance bulletin from NIC-http://informatics.nic.in/try_dispnews.asp ?newsid=22 & module_ number=apr_4

About the Authors

Siddhartha Ghosh, Siddhartha Ghosh is an Associate Professor in the Dept. of CSE, GNITS, Hyderabad. He has Completed B.Tech. from TEC, Agartala presently NIT Agartala, Tripura He has an M.Tech. degree from University of Hyderabad. He is perusing Ph.D. from Osmania University, Hyderabad. He has eight years of Teaching experience and two years Industry experience. He has 15 national and international publications. He is member of CSI, ISTE and IE. He has vast experience in NBA accreditation work and ISO 9000:2001 certifications for educational institutes. His areas of interest are Artificial Intelligence, Natural Language Processing, Machine Learning, Web Technologies and Indian Languages.

Sameen S Fatima, Dr. S Sameen Fatima is a Professor in the Department of Computer Science and Engineering at Osmania University. She is currently on a sabbatical at BITS Pilani - Dubai. Her research interests include Information Retrieval Systems, Computational Linguistics and Text Mining. She received her B.Tech. from Jawaharlal Nehru Technological University in Electronics and Communication Engineering and got her M.Phil. in Computer Methods from Hyderabad University. She joined the Department of Computer Science and Engineering, Osmania University as a faculty member in 1984 in its formative years. During 1989-94, she received her M.S. in Computer Science and worked at the University of Massachusetts at Amherst, USA