Application of Natural Language Processing Techniques in e-Governance

M A Moni

Abstract

Natural Language Processing (NLP) is the technique of engineering our language through machine (the computer) by which we can overcome the language barrier and the difference between man and machine. On the other hand, 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 (ICT) systems to deliver services through multiple channels that are accessible, fast, secure, reliable, seamless and coherent. To implement better G2G (government-to-government), G2B (government-to-business), G2E (government-to-enterprise) and G2C (government-to-citizen) services we should not only utilize ICT, we also have to be serious about implementing Natural Language Processing (NLP) techniques to reach the masses and make e-Governance successful. This paper shows the need for applying NLP technologies in the field of e-Governance and also tries to focus on the issues, which can be resolved easily with the help of these modern technologies. It also shows the advantages of applying NLP in e-Governance.

Introduction

Good governance is a process by which individuals and institutions, public and private, manage their common affairs. It refers to the interaction between the public sector and civil society for collective decision making (Kazancigil, 1998). Good governance has eight major characteristics: it is participatory, consensus oriented, accountable, transparent, responsive, effective and efficient, equitable and inclusive, and follows the rule of law. It assures that corruption is minimized, the views of minorities are taken into account, and that the voices of the most vulnerable in society are heard in decision making. It is also responsive to the present and future needs of society.

E-Governance is a strategic issue that relates to re-designing national administrative processes at the policy-making level to make government more transient, efficient and service-oriented. E-Governance should be seen to encompass all ICTs, but the key innovation is computer networks – from intranets to the Internet – creating a wide range of new digital connections:

- Within government – permitting “joined-up thinking”
- Between government and NGOs/citizens – strengthening accountability
- Between government and business/citizens – transforming service delivery
- Within and between NGOs – supporting learning and concerted action
- Within and between communities – building social and economic development

The novelty of using new technologies in governance is that it expands beyond internal government operations to include electronic service delivery to the public and the subsequent interaction between the citizen and the government. This potential for
Proceedings, 8th International Conference on Language and Development, Bangladesh 2009

interactivity can be identified as one of the most important elements in the way e-Governance will change the nature of government (Heeks, 2001). The major aims of e-Governance are to:

- Strengthen and upgrade the present infrastructure of information technology in the ministry
- Bring in transparency by placing information, to the extent possible, in the public domain
- Link the functioning of different offices to bring about synergy and enable effective and timely decision making
- Develop public grievance and complaint systems and eliminate the need for personal visits to offices by the public
- Make available forms and information on the Internet and the facility of filling the applications from remote rural areas (web-enabled application systems)

The next major point should then be:

- Application of modern technologies by e-Governance (i.e., NLP techniques).

According to Wikipedia, NLP 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.

**Major Tasks in NLP**

**Speech synthesis:** Although this may not at first sight appear “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.

**Information retrieval (IR):** Concerned with storing, searching and retrieving information. It is a separate field within computer science (closer to databases), but IR relies on some NLP methods (for example, stemming). Some current research and applications seek to bridge the gap between IR and NLP.

**Machine translation:** Automatically translating from one human language to another.

**Speech recognition:** Given a sound clip of a person or people speaking, the task of producing a text dictation of the speaker(s). (The opposite of text to speech.)
Some others tasks in NLP include:

- Automatic summarization
- Foreign language reading aid
- Foreign language writing aid
- Optical character recognition
- Text to speech translation
- Text-proofing
- Spell checkers
- Grammar checkers
- Online dictionaries
- Named entity recognition
- Question answering
- Spoken dialogue management
- Text simplification

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 into the major e-Governance projects for any country.

**Major e-Governance Related Activities**

The overall architecture for e-Governance needs to ensure that the architecture components are extensible and scalable to adapt to the changing environments. The e-Governance applications that are emerging as islands of successes have to be inter-operable.

*Country Portal*

Country portal is a one-stop shop for all government business in any country (example in Figure 1). There are various government 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 and sub-governments.

Application of NLP techniques in e-Governance services offered under this category are:

- 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
State Portals

State portals are the single-window government entry points of various states. They incorporate most of the same features mentioned above as well as links to central government resources.

EDP

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, such as:

- Information about laws, regulations, codes and procedures involved in running businesses in the country
- Filing of taxation, returns, claims and refunds
- Registration and start-up of a company, business, SME or any form of organization
- Issue of licenses, permits and certification
- Renewal of licenses, permits and certification
- 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 and other regulatory bodies
- Recruitment of skilled manpower
- Forms required for various services
• Feedback on various initiatives and discussion forums
• Source of funding (banking)
• Obtaining land and building approvals
• Compliance with 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)
• Act as the interface to international agencies

**e-Procurement**

Each government department and agency at national and state levels procures material. With the use of ICT in procurement processes, the overhead cost will go down. There are examples in the West where governments have reduced their costs by over 50% by using e-Procurement.

**NLP Techniques in e-Governance**

NLP-based e-Governance systems are already in use in some countries. For example, the Natural Language Based Information Retrieval System Anveshak (The Quester) can efficiently and accurately provide explicit information in natural language text to questions on a certain document. More initiative has to be taken in various levels of state and central government to implement NLP-based e-Governance models or systems in every country. When it is matter of multilingual countries like India then NLP may be the black horse of e-Governance.

E-Governance is not only for educated people but for everyone. From government employees to farmers in villages, anyone can use the facilities offered by e-Governance systems. For example, a farmer wants to post the price for his rice to sell through the kiosks available in a village. If he is an educated person he can easily use the kiosk or Internet-connected computer which is kept in the Gram Panchayet office. If he is not educated, then a speech recognition (speech to text converter) system may work for him. If the speech recognition system is also enhanced with a speech translation system then even though the person may use his mother language like Bengali, the price quote can be in English. The NLP-based machine translation system will be used here as a human friend. Blind people can use speech recognition systems easily. They can fill forms for any government job application, death and birth registration, or online application for including their name in a voter list. They can just utter the sentences and words and the system will help them to feed it to 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.

**Use of Natural Language in Government Websites**

As government websites are full of information, forms and data, they are the major vehicles for offering e-Governance services. In this context the use of multilingual web portals can change the face of any government as well as help in overall development of any country or state. In Bangladesh many NGOs and government agencies offer multilingual websites or local language based websites. In this regard we can name few of them which have already become popular for their services. Natural language is used in the different websites shown in Figures 2, 3 and 4.

![Figure 2: National website of Bangladesh government](http://www.webbangladesh.com/)

Figure 3: Website of Web Bangladesh (http://www.webbangladesh.com/)
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 linguistics, psychology, engineering and computer science expertise so that computer programs can be made to produce high-quality natural language text from computer-internal representations of information.

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, converting sentence-sized chunks of representation into grammatically correct sentences.

Any e-Governance system which uses natural language generation techniques, can provide better service. For example, a natural language generation based e-Governance system can help people to listen to the text as audio files once any one selects them. Even though the person may not be familiar with that text he can understand and speak it. Blind people can hear news, information, processes, and job vacancy information. Elderly people with poor eyesight can use this tool. The NLP-based question and answer systems can help people with anything. A farmer who cannot read text can even get help from a website, like how to use fertilizer in the field, grow new crops or water them using new techniques. Government institutions like Bangladesh Railway and Bangladesh Air can use such systems for common people to make the services much people friendly.

The official website of the First Citizen of Bangladesh 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 Honoarble Prime Minister and the common citizens.

Concluding Remarks

Moving house, having a baby, setting up a business; there are numerous reasons why citizens or companies contact public administrations, and local governments are under increasing pressure to improve the quality and speed of service delivery. More than offering just web-based information, the time has come to give citizens personalized assistance and solutions to enquiries, entered entirely using natural language processing techniques. It is not a new concept and the time has come for more R&D work to be initiated at government level to take up this issue more seriously and make e-Governance a common man’s friend.

References

Baweja, Y 2005 Does E-Governance Mean Good Governance? Yojana 49.


E-Governance Bulletin from NIC. http://informatics.nic.in/try_dispnews.asp ?newsid=22 & module_number=apr_4


Ghosh, S and Fatima, S S 2007 Use of Local Languages Portals. CSI Communication, 4-12.


Standards and Architecture for E-Governance Applications http://egovstandards.gov.in/localisation_wg/localisation_std