The Machine's Role in Human's Service Automation and Knowledge Sharing

Mihály Héder

MTA SZTAKI Institute for Computer Science and Control
Hungarian Academy of Sciences / Department of Philosophy and History of Science, Budapest University of Technology and Economics
Kende u. 13-17. Budapest, H-1111

Tel: +36 1 279 6027

Fax: +36 1 279 6021

mihaly.heder@sztaki.mta.hu
Originally Presented at SWIIS 2012, Waterford, Ireland, June 11-13, 2012.

The possibility of interacting with remote services in natural language opens up new opportunities for sharing knowledge and for automating services. Easy-to-use, text-based interfaces might provide more democratic access to legal information, government services, and everyday knowledge as well. However, the methodology of engineering robust natural language interfaces is very diverse and widely deployed solutions are still yet to come. The main contribution is a detailed problem analysis on the theoretical level, which reveals that a text-based interface is best understood as an artificial agent that represents the interests of the remote party who is separated in time and space from the client. A possible ethical issue about the development of such an agent is also discussed.

Natural Languages, Interactive Approaches, Communication Networks, Artificial Intelligence, Software Specification

Introduction

The focus of this paper is on the automation of certain services that are accessed in natural language by the human end-users. This includes a multitude of tasks: government services, information sharing and gathering, front-desk operations, a company's customer services, commerce, and more.

As Rothstein and Teorell (2008) points out, in developing and transition countries that are often less stable, the presence of impartial government institutions are very important enabling factors of progress. They argue that the former feudalistic relations between government officials and residents persist in cognitive structures even after e.g. legal reforms were formally carried out. That is, residents do not expect impartial decisions from government officials, who indeed might have problems acquiring the idea of impartial public administration. The lack of trust in an impartial government creates uncertainty, which keeps residents back from investing time and resources that are required for achieving long term goals. On a macro scale this problem hinders growth and social welfare.

In our digital age many of the aforementioned public and commercial services can be made accessible for an end-user who is equipped with a suitable appliance – a desktop, laptop, mobile phone, tablet, etc. The electronic way is usually more economic for both parties; thus, end-users are often incentivized to use electronic services rather than having to go to a physical office in person. This phenomena has many benefits: services can operate 24/7 thus require less adaptation from the users and saving time; well-designed systems eliminate waiting in queue, thus saving time; and there is no need to travel, which also saves money and time.

Moreover, residents might expect more impartiality from a computerized interface to government services, than from the local persons in charge – just because they will be dealing with a computer that does not have its personal agenda. In a naive utopistic vision giving the decision making role to computers instead of humans might indeed result in more impartial and disinterested decisions. Even if we are less optimistic, we can still expect more transparency and more equal standards from computerized public administration. So, at the first glance government service automation appears as a great tool for facilitating stability and growth.

There are problems however. The lack of ubiquitous computer and Internet access is a serious issue that needs to be addressed, but is out of the scope of this paper. In some domains safe authentication and unnecessary keeping of personal data is also problematic; these are not discussed here.

In the focus of this article, there is another serious problem: most of the electronic interfaces of the services work in quite a different way than their regular counterparts (e.g. real life offices). These interfaces usually reflect the conceptual framework of the software developers, which is not equally intelligible for everyone – this is why the more technical-minded can benefit more from these services than others.

It is an undeniable tendency that these interfaces are getting more and more intuitive as software developers are reflecting to user feedback. Also there are signs that the interfaces are getting ever more uniform, as best practices are spreading, which helps the users to use their experience with one service when using another.

In certain domains, administration involves producing textual documents like applications, complaint letters, legal questions, information requests, etc. In the majority of these cases electronic interfaces are only used as communication means only and processing of the content remains a human actor's task. In this paper the possibility of automatic processing and reply is discussed in those cases where user input is plain text rather than form data. Automatic or semi-automatic language processing systems might prove useful as an alternative to form-based electronic administration for those who are less comfortable with forms and more with natural language.

Natural language processing is an established field of research (for a good textbook see Jurafsky and Martin, 2008). However, the application and efficiency of its results largely depends on the practical context of the problem.

The aim of this paper is to take a step back and attempt to make an analysis of the pragmatics of the text-based service automation problem in general. This includes investigating both the preliminaries and some ethical consequences of the problem. I think that these issues must be addressed before service automation can be useul as a real enabler in unstable or developing regions instead of something the widens the gap between those who are familiar with computer systems and those who aren't.

The Message Representation Creation Problem

To begin the investigation of the problem, it will prove useful to give a very selective account of the history of machine-mediated communication.

The invention of written language provided a way to separate the sender of the message and the receiver in both time and space.

However, initially message transport was quite slow. This was changed by semaphores (Holzmann and Pehrson, 1994) and later by electronic and radio communication. These latter technologies made it possible to be separated in space but still communicate in real time.

With introducing the computer, as an intermediate party new opportunities opened up (for more on this aspect, see Vámos 2010). It became possible to separate the structure of the messages from representation, thus accessing messages with many different devices and interfaces. It also became possible to programmatically execute an administration workflow with conditionals that depend on user messages. The meanings of the messages are quite clear if the system uses fill-out forms, but less so when only texts are available.

Figure 1 illustrates that as we add a computer as an active intermediate party, now there are two interfaces of communication. Let us investigate what this means if this is a system that works with natural language.

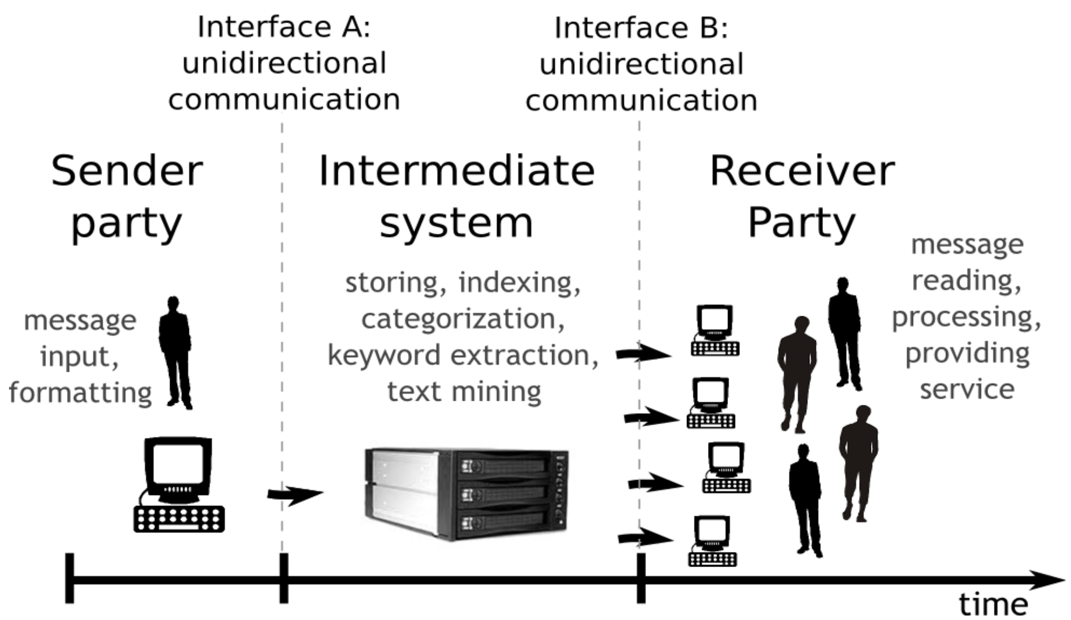

Figure 1: An overview of computerized communication

Interface A is between the end user and the computer. It is a real-time communication, which, however is unidirectional (see the next section) as the computer in general is not able to understand the user's messages and its reaction is only about the transport of the message (e.g. status of the transfer, etc).

Interface B is between the computer system and the human recipient of the message. It is also a real-time, unidirectional channel when we talk about meaningful natural language communication. Naturally, other types of interaction, e.g. keyword-based search or message browsing is commonplace.

For example, in the case of an electronic complaint letter, the user writes the letter on *interface A*, the computer system stores it along with a set of metadata (e.g., writing date, etc.), and then the message is delivered to an administrative person through *interface B*. A big difference between form-based systems and text-based systems is that in form-based systems *interface B* might be missing or is less frequently used, as the computer system might be able to handle the communication and resolve the case according to a pre-programmed policy through *interface A*. This is possible because the input data and its meaning are pre-defined.

However, handling a free-form, text-based message according to a pre-programmed policy would involve some kind of understanding of the human language, which problem is the subject of the current article.

As a pre-programmed policy or work flow needs to operate on structured input data with fixed data types and meaning, in this scenario we can define natural language understanding as a successful mapping or extraction of the information contained in the text to structures that are operational for a program code.

This means that a machine representation must be assigned to the message as a whole (and not sentence-by-sentence for example) and this representation is also a function of the program that will operate upon it to provide a service. Ultimately, the proper representation changes from problem to problem, and one piece of text might have several representations.

Let us define the task of creating the problem-specific structured machine representation of natural text as the *Message Representation Creation* (MRC) problem.

Solving an MRC problem might involve named entity recognition, sentence analysis, or part of speech analysis, but also simple categorization based on word statistics. The way of solving MRC problem always depends on the details of the task.

There is one important aspect however, that is essential for the robustness of all systems tackling MRC: the mode of communication on *interface A*.

Bidirectional communication in Message Representation Creation

As MRC involves a practical understanding[1] of natural language messages, an analysis of the communication context can be relevant. Figure 2 and Figure 3 explain the difference between unidirectional and bidirectional communication. In general, unidirectional communication is less robust because there is no instant possibility to disambiguate or ask for clarification on the recipient's part. This is common in the task of writing a book, delivering a speech to a large audience, and even in writing a letter. Communication has to be unidirectional because of separation in time and space, or because there is one sender and multiple receivers, and also in cases when the receiver only transmits but does not understand language, as it is the case with interface A explained above.

[1] „Understanding" used in a limited sense here, as the problematic question of what is „real" understanding is not discussed.

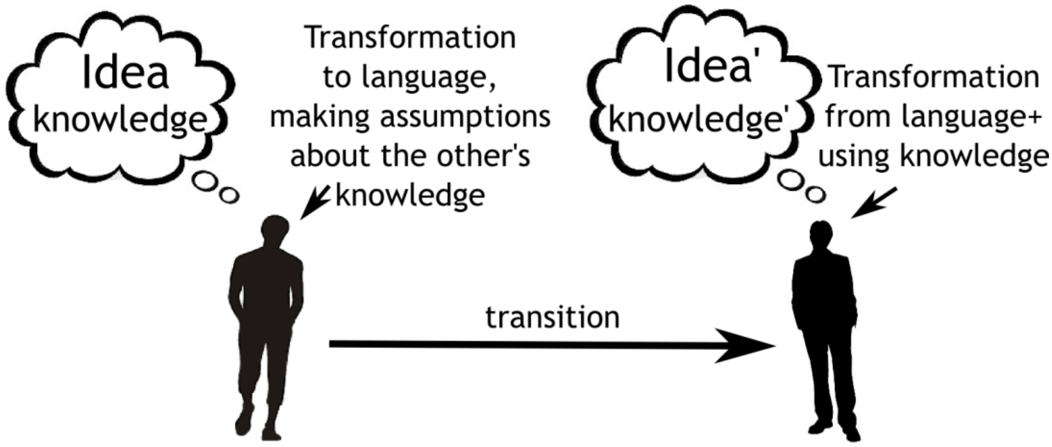

Figure 2. Unidirectional communication

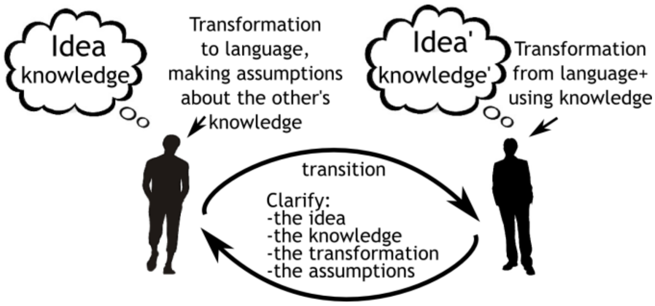

Figure 3. Bidirectional communication

In the case of bidirectional communication there are more ways to ensure the understanding of the message. The receiver has the capability to decide whether the message is consistent for him or her. This capability not only signals the inconsistency but also tells what is missing or ambiguous in the current understanding of the message that prevents it from being consistent. In general, the availability of a bidirectional channel makes communication less demanding for both the message sender and recipient party.

There are two other important factors that make bidirectional communication necessary besides the fact that it is less demanding. First, the proper and fully detailed representation of the text message might be still insufficient. That is, even though we understand every bit of the message, there might be additional information needs to proceed while providing the service that is simply omitted by the writer. For example in the complaint domain we have made experiments with (see later) it is common that the person who fills the complaint simply does not specify crucial bits of data, e.g. the type of pension she/he is receiving, or even the name or the branch of the bank that is the subject of the complaint. Naturally, in this case even the best human readers would have to ask additional questions from the sender of the message. Now, if we replace the human with a computer on *Interface A*, then gathering additional data that is not in the text has to be as much important as extraction of information from the given text.

Second, there is the question of the elapsed time between writing and processing. Texts are often thought of as they had a fixed, static meaning that needs to be discovered either automatically or manually. Computer Scientists have to be reminded that that is not true in general as meaning nearly always depends on context. There might be texts where the context is negligible but in many cases it is not. For example if we have a complaint letter to process from 2006 now in 2013, and there are references to the “city major” or certain paragraphs from legal documents then we need to somehow recall the name of the city major and the legislation that has changed since. This is not impossible, but adds a huge additional demand against the knowledge base we use. On the other hand, if we resolved those at the time of writing the problem would be much smaller. A much less acute problem is that even the meaning and use of words can change over time – this is not an issue yet, but will be for e.g. the projects that aim to archive the internet.

However, natural language processing methods are not optimized for this kind of communication. This is because their primary function is to process large text corpora like crawled web pages, encyclopedias, or similar databases. In the case of a web search engine there is no way to involve human assistance while processing the texts, so the processing has to take place fully automatically. This class of problem shapes how the majority of NLP tools approaches text processing. It is still quite hard to find NLP tools that can use the human in the loop for information extraction.

With the MRC problem however, there is no reason to attempt the communication in a fully automatic way. Technically it would be no problem to reach the user with questions: there is constant internet connectivity, and having *html5*, *websocket*, *AJAX*, etc. all the necessary the tools are ready. The intermediate system could act as an agent (Russel and Norvig 2010, chapter 2), which *represents the receiver's information needs and acts proactively*. This involves knowledge about the range of the required information, the capability to extract that information from the text, and also a measure of consistency. The consistency measure is used to determine whether additional questions need to be asked and what those questions are.

This mode of operation has an effect on the particular methods used for text processing. For example, when using decision tree based categorization, it is straightforward how to generate questions: "when in doubt at some point down in the tree, ask a multiple choice question". It is much less easy to do the same in the case of a neural net based solution.

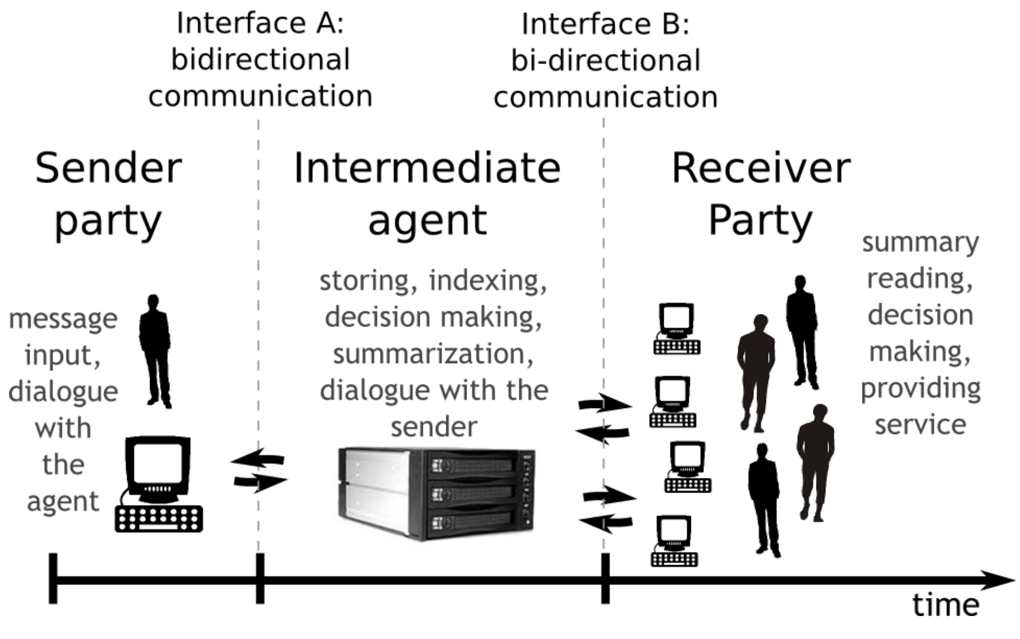

Figure 4. The intermediate system as an agent representing the receiver party's communication needs

We can see that it is not rational to rely on unidirectional communication only when solving the MRC problem. A much more suitable design relies on bidirectional communication on both A and B interfaces, as depicted in Figure 4.

In this architecture the system acts as a surrogate of the receiver – it should ensure that the message contains all required information for providing the service, and also that the sender is served in a polite manner throughout the dialog that might be initiated by the system.

Prototypes and Applications

At MTA Sztaki we have developed two systems that are based on the principles detailed above.

The first is a complaint letter-writing interface, which was tested with approximately 900 Hungarian complaint letters written to the Ministry of Justice, and 7000 Bank Complaints written to the website www.complaintsboard.com. (Héder et. al. 2011b; Vámos and Héder 2011)

This system provides a what-you-see-is-what-you-get type of web editor interface, through which the sender can compose his or her personal message. The system has knowledge about the most regular case types. The set of these model cases is acquired from a learning corpus with the involvement of experts.

One way of the acquisition is to ask a few experts to create clusters from the complaint letters and then to identify the varying features of the cases and separate them from the fixed features.

Another way might be taken if we have a large number of human participants with less knowledge of the specific domain: we can have the participants write 5 or more short summaries of the letters, about 3-5 sentences in each summary. We have conducted this experiment with approximately 200 students who all had to write summaries for 3 letters. There were 100 letters in the pool, which means that we had 6 summaries per letter.

The experiments revealed that the summaries are very similar for the majority of the letters but more diverse for others. We could use this information to tell apart the easily and consensually understandable letters from the confusing ones.

Also, the summaries of different letters are more similar to each other than the original letters, as the summaries contain the relevant information and omit the irrelevant. This gives an opportunity to abstract away from the letters to a small number of model cases.

With these methods we acquired 20-40 abstract cases for each corpus. The cases were represented in a frame format that is a combination of Minksy's (Minsky 1975) and Fillmore's (Fillmore 1976; Baker et. al. 1998) frame idea. For every frame there are *evokers*, which are keywords or phrases that activate the given frames. There are also *slots*, which need to be filled with case-specific information.

Given the low number of frames we could manually create question templates in human language for each.

For example, in the bank complaint corpus a typical case is STAFF PROBLEM, which is evoked by certain phrases like "harassed," "they did not even bother," "unpleasant," etc. If the system encounters one or more of these expressions it activates the STAFF PROBLEM frame. For the STAFF PROBLEM frame there are slots like the name of the BANK REPRESENTATIVES involved, the TYPE OF THE PROBLEM, the BRANCH NAME, the DATE, etc. We also have information about what kind of data usually fills these slots. From this information we can formulate natural language questions that ask for missing data and for confirmation of the system's hypothesis about the specific STAFF PROBLEM case. In other words, the system initiates a dialog in order to help its own understanding of the letter. Once the frame instances are completed, the system has a machine representation of the message, and is therefore able to execute a pre-programmed workflow according to it. This system is right now in a demonstrable pilot phase, unfortunately without end-user application yet.

Another system is Sztakipedia[2] (Héder et. al. 2011a, Héder 2011), which is used in production by a growing number of Wikipedia editors. This system is an extension for editing interface of the popular free encyclopedia. Meeting the quality requirements posed by the Wiki community can be hard to meet for the new editors. The quality requirements against a newly written article include a

[2] *http://pedia.sztaki.hu*

proper amount of links to other Wiki pages, good categorization, filled infoboxes (structural data that can be found on many Wiki pages, e.g. cities), and bibliography references. Sztakipedia as an agent is able to represent these quality requirements of the community while communicating with the sender of the message, who is the new editor in this case. By offering a link, infobox, category and related literature suggestions it helps the sender enhance the quality of his or her page.

Sztakipedia is integrated with DBpedia Spotlight (Mendes et.al. 2011), which in turn acquires its knowledge from Wikipedia itself. This enables a virtuous cycle of knowledge creation as the semantics in existing Wikipedia articles facilitate the creation of new ones.

An Ethical Issue

The capability of a machine to maintain a dialog like a human might change the everyday user's understanding of intelligence in general. As this phenomenon might degrade or de-humanize certain professions in the eyes of some end-users, we have to investigate it as a potential ethical issue of designing natural language interfaces. This issue is of course only an additional one to the well known ethical issues concerning autonomous machine decision making, job losses because of automation, the ethics of safety and error, etc. (see Lin et. al. 2011; Anderson and Anderson 2011)

Figure 5 explains the issue in detail. Let us consider the usual case in which a human representative serves a client by phone or by a client chat over the network. In this case the perception of the representative is consistent: it acts like a human and is indeed a human.

However, if natural language agents become widely deployed, that perception might become inconsistent, as the client will be informed that she or he is interacting with a machine, but the interaction itself will be similar like with a human. This inconsistent perception might be tacit for the person.

This inconsistency might lead to two different kinds of attribution errors. First, one might attribute other human features to a machine such as emotions, mood, "soul," etc. There are weak signs of this phenomenon among everyday computer users.

However, there is another kind of attribution error that can originate in the inconsistent perception. A human might attribute machine features to other humans, tacitly thinking things like "telephone operators are biorobots" or "there is nothing special in chess playing, one has to calculate a lot." This is what we can call a dehumanizing perception of tasks that require intelligence.

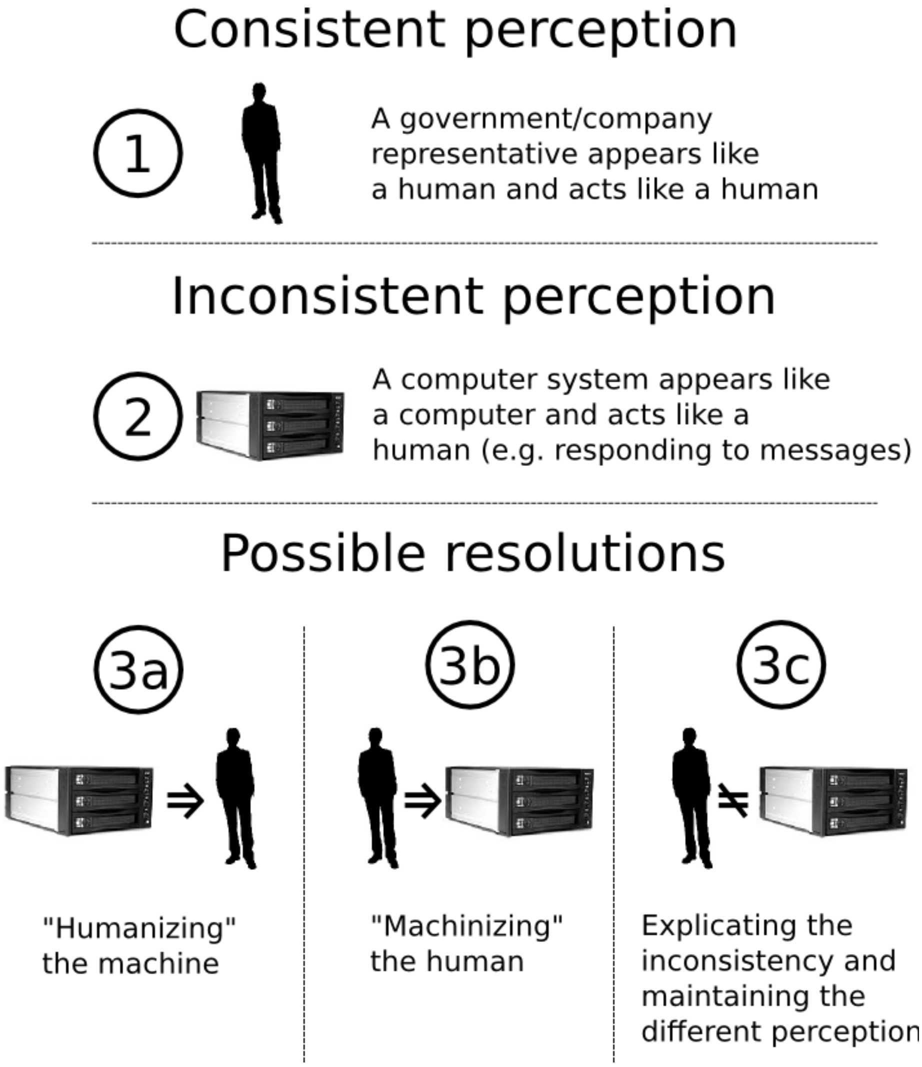

Figure 5. An overview of the potential problem of inconsistent perception

Correcting spelling errors and playing various games are the first areas of intelligence where computers have made widespread success. As a consequence, the common understanding of the nature of the capabilities that are required to carry out these tasks has changed. The value that someone might assign to the success in these tasks has potentially degraded.

There are two ways out from this attribution error. The first is to make a machine's appearance as similar to a human as possible, thus reducing the inconsistency of the perception. This would involve creating a friendly human look and other features for the machines, as convincing as possible.

The other way is to explicate the different nature of the machine agent. This means that the invention of new modes of interaction is required that are very distinct from human interaction, but still able to carry out the task of bidirectional communication.

Conclusions

On the EmTech 2012 conference Nicholas Negroponte reported about a rather radical experiment in which they gave tablet computers to illiterate children in Etiopia without any further instructions or assistance[3]. They have learnt that the children were able to figure out how to use the applications rather quickly. This indicates that in the long term, provided that people everywhere have access to cheap computing devices we can hope that using computerized services wont be a big problem for anyone.

However, important public administration or commercial services are for the adults. By establishing automated on-line services, one can provide cheaper and faster service, that, in theory helps residents save time and effort. But if the user interfaces of those services are merely reflecting the database tables they should populate, then only those will benefit from them who already have experience with computers – in other words, it won't be inclusive enough, they can become and obstacle, further widening the gap between different social groups and thus generating instability.

Therefore, the author thinks that large emphasis should be put on the user interface, and natural language understanding should play a great part in it. In the paper the theoretical background of creating services that are able to communicate in natural language was discussed with a philosophical bent.

First, a precise analysis of the communication situation was attempted. As a result, the Message Representation Creation (MRC) problem was defined, in which the goal is to gather machine-readable information that can be processed by a pre-programmed code that changes from task-to-task. As there is a real-time interface between the client and the system, it is rational to attempt to get the information in bidirectional communication, by initiating dialogs.

A demonstrable complaint letter processing system and Sztakipedia, an assistant for Wikipedia that is in production, was shortly discussed. These solutions are however still limited in various ways. One limitation is in the degree of automation. These systems might extract entities, relations and other important information from the text, in social issues judgement and decision making will always be necessary. Another limitation of the system is its language dependence. And finally, the major barriers of adoption of systems like this are the missing legislation and motivation on the government's side.

Finally, the crucial ethical issue of the *erroneous attribution of machine features to humans* was presented. Among the many ethical problems with machine automation, this is less obvious and undiscussed, but is still a serious matter.

Many people are doing jobs for a living that can be potentially automatized. I believe that in the future these will include jobs that involve natural language understanding. People tend to associate a rather low value to those activities that can be done by machines – which reveals a functionalist position by the way. When natural language tasks will be in this circle, the territory of the "real" human intellect – as perceived from this functionalist point of view – will shrink to tasks that involve creativity, improvisation, and maybe emotions, that are not attributed to computers yet.

This can lead to serious tensions in societies, because the majority will either do jobs that could be done by a computer, or actually become unemployed because of computers. While the latter problem is well known from the age of industrialization, the former is novel one, that can be addressed by giving a characteristically non-human appearance to intelligent systems.

[3] http://www2.technologyreview.com/emtech/12/video/#!/watch/nicholas-negroponte-teaching-children-to-learn/

Acknowledgement

The research was supported by MTA SZTAKI, the grants TÁMOP - 4.2.2.B-10/1--2010-0009 and OTKA K72598

References

Anderson, M and Anderson S. L. (2011) *Machine Ethics*. Cambridge University Press. ISBN: 0521112354.

Baker, C.F, Fillmore, C.J., Lowe, J.B. (1998) The Berkeley FrameNet project. In: *Proceedings of the 17th International Conference on Computational Linguistics*, volume 1, p. 86-90. Montreal.

Fillmore, C.J. (1976). Frame Semantics and the nature of language. In: *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280:20-32.

Héder M., Farkas T., Oláh T., and Solt I. (2011a). Sztakipedia – Mashing Up Natural Language Processing, Recommender Systems and Search Engines to Support Wiki Article Editing. *Demo paper for ESWC 11 AI Mashup Contest*, Heraklion.

Héder M. (2011). Integrating artificial intelligence solutions into interfaces of online knowledge production. In *ICIC Express Letters*, Volume 5, Issue 12, Pages 4395-4401.

Héder M, Bártházi E, Vámos T (2011b). Natural Language Understanding in Governance Service Automation. In: *Proceedings of the IFAC 18th World Congress*, Milan.

Holzmann, G. J. and Pehrson, B. (1994), *The early history of data networks*, Wiley-IEEE Computer Society Press. ISBN:0818667826.

Jurafsky, D. and Martin, J.H. (2008) *Speech and Language Processing* (2nd Edition). Pearson Prentice Hall, ISBN: 0131873210.

Lin, P., Abney K., Bekey G. A. (2011) *Robot Ethics: The Ethical and Social Implications of Robotics*. The MIT Press. ISBN: 0262016664.

Mendes, P., Jakob M., García-Silva A. and Bizer C. (2011) DBpedia Spotlight: Shedding Light on the Web of Documents. In the *Proceedings of the 7th International Conference on Semantic Systems (I-Semantics)*. Graz, Austria, 7–9 September 2011.

Minsky M. (1974) A framework for representing knowledge. Technical report, Cambridge, MA.

Rothstein, B, Teorell, J. (2008). What is Quality of Government? A Theory of Impartial Government Institutions. In *Governance: an International Journal of Policy, Administration and Institutions.* Vol. 21. No. 2. April 2008. Pages 165-190.

Russell, S.J, Norvig, P. (2010), *Artificial Intelligence – A modern Approach*. Chapter 2. Pearson Education, Upper Saddle River, New Jersey.

Vámos, T. (2010) *Knowledge and Computing*. Central European University Press, Budapest-New York. ISBN: 9639776645

Vámos, T. and Héder, M. (2011) Understanding in distributed social-administrative network systems. In: *MIPRO 2011 – 34th International Convention on Information and Communication Technology, Electronics and Microelectronics - Proceedings*. Opatija.