LTC 2007, Poznan: Panel discussion on Homeland Security. Chairman's Statement

Dafydd Gibbon
Fakultät für Linguistik und Literaturwissenschaft
Universität Bielefeld
Postfach 100131, D-33739 Bielefeld
gibbon@uni-bielefeld.de

Abstract

The statement of the panel discussion chairmen is presented. The invited panelists who confirmed their participation before this text went under press are: Dafydd Gibbon (Uni Bielefeld) - moderator, Bernardo Magnini (ITC irst, Trento, Italy), Keith Miller (MITRA, USA), Kimmo Rossi (INFSO, EC, Luxembourg), Emil Pływaczewski (University in Białystok, Poland), Sebastian Serwiak (ESRIF, Bruxelles), Zygmunt Vetulani (UAM, Poland). The Chairman's statement is appended by position statements provided by the invited panelists.

A linguist could approach the security topic by asking what "security" means and what it implies. Apparently, security is located on a scale between the hazardous extreme of "no risk, no fun", on the one hand, and an unattainable total freedom from risk on the other. It is also clear that the attainment of security requires social constraints on interaction which limit individual freedom in favour of the survival of agreed social and political values, but which limit tendencies to totalitarian control. For discussion of language and information technologies, some key ethical issues are:

- self-determination of personal information by individuals in contrast to stealthy access to information by executive governmental bodies,
- determination of which parties have freedom of access to personal information,
- determination of which information is collated into overall profiles ("the glass citizen").

Behind these issues are the motivations for security, which are perceived or apparent threats, and the relevant institutions by which security, once decided on, is realised, depending on the type of threat. The following is a small and informal selection:

- natural or human influenced disaster: insurance, emergency services (police, fire, medical, repair services)
- financial: insurance, underwriting,
- property crime: policing,
- terrorism: diplomatic and military.

Finally there are the technological issues involved in implementing security. The present workshop touches on the role of the human language technologies in a subset of these areas, in which it turns out that security issues cover a very broad range of infrastructural areas which already involve the deployment of human language technologies for other reasons:

1. Communication functionality and Human Language Technologies:
   1. Adherence to legal and ethical standards.
   2. Robust behaviour in adverse environments.
   3. Multilingual systems (translation, summarisation).
   4. Bilateral and multi-party dialogue management, including dictation.
   5. Language understanding.
   6. Efficient browsing and general search of language and speech databases.
   7. Identification / verification of language and speaker identities.
2. Structure and physical form of Human Language Technologies:
   1. Data resources:
      1. Large, information-rich machine readable lexicons (lexical databases).
      2. Large tagged corpora for text (largely automatisable).
      3. Large tagged corpora for speech (partly automatisable).
   2. Language (text) processing systems:
      1. Parsing: Shallow parsing of arbitrary text.
      2. Generation: Report generation; dictation output.
      3. Translation: Translation memory and terminology tools for translation support.
   3. Spoken language (speech) processing systems:
      1. Speech recognition: dictation input.
      2. Speech synthesis: information provision in acoustically hostile environments.
      3. Speaker verification and identification.

4. Multimodal systems:
    1. Interfaces between speech and language systems.
    2. Speech related gesture processing (analysis and avatar synthesis).

Discussion on any or all of these issues is very welcome, both from the panelists and from the audience.

Bernardo Magnini
ITC irst, Trento, Italy (magnini@itc.it)

I will bring into the panel the perspective of two related areas where text processing technologies play a crucial role for homeland security. In Speech Analytics information extraction techniques are used to process large amounts of automatic transcriptions from audio/video data in order to mine them for relevant information. In Open Domain Question Answering the effort is to provide intelligent analysts with tools for fast and reliable access to large collections of textual information, in multiple languages, according to very with specific user needs.

Keith Miller
The MITRE Corp., USA(keith@mitre.org)

The topic of "Homeland Security" is sufficiently broad that one could imagine relevant applications for practically any application from the spectrum of Human Language Technologies. It is possible to imagine potential uses for everything from information retrieval, through information extraction (including entities, events, relationships, etc.) and even speech recognition, to current research in intent recognition. Add to this the indisputable fact that not all relevant information is in English, and one can certainly add machine translation (and its companion technologies) to the list. Finally, taking into account that it is desirable to do as much processing "in language" -- that is in the original language of the text – as possible, and there is an argument that Homeland Security applications could make use of the range of above-mentioned technologies in a wide variety of languages.

Kimmo Rossi
EC, INFSO, Luxemburg
(kimmo.rossi@ec.europa.eu)

In security-related language technology applications, it is very important to ensure conformity with appropriate legislation on privacy and on the treatment of personal data. Ethical implicatons of technology need to be assessed beforehand, involving experts with the appropriate competence (legal, political, technical, organisational etc.). In European-funded research (FP7, for example), these considerations have been extensively prescribed and the ethical issues need to be documented and addressed. Any project should be designed in transparent terms and stand the scrutiny of the media, politicians and demonstrate legal compliance.

Sebastian Serwiak
ESRIF (Brussel)
Of particular interest for operational practice are processing technologies assuming the voice input:
- automatic identification of speakers as an instrument to be applied in operational practice (by police and similar services),
- automatic speech processing and in particular speech-to-text conversion followed by text understanding are crucial elements of advanced systems involving automatic understanding.

Zygmunt Vetulani
UAM, Poznań, Poland (vetulani@amu.edu.pl)

Homeland Security is a challenge for HLT's for many reasons:
-HLT's based applications we intend to develop must be highly performant, human friendly, robust, safe, efficient... This means that expectations are extremly high.
- both technology development and applications must respect law, stick to high moral standards,
- real world scale HS problems require technical advancement which is in many respects much beyond the current state-of-the-art (speech-to-text conversion, text understanding)
- multicultural and international nature of many HS problems makes that the HLT tools and methods involved must also have multinational, multicultural and multilingual character.
- critical issue: technical advancements of tools, methodplogies and resources is not the same for all languages.
- many technological gaps are to be completed for different languages. It is important to keep under control the HLT developent and reduce many-speed development.