Proposal: TEI-SIG "Computer-Mediated Communication"

Context: In the past three decades, computer networks and especially the internet have brought forth new and emerging genres of interpersonal communication (*computer-mediated communication*, henceforth *CMC*). Even though there's been a lot of research on cmc genres and on language use on the internet in linguistics and social sciences as well as in the field of natural language processing, there are still no common standards for the representation and annotation of these new forms of communication and their structural and linguistic peculiarities. Being able to represent cmc data on the basis of an encoding framework such as the TEI which is broadly acknowledged within the field of digital humanities would allow for an interchange of data between research groups and for building interoperable cmc corpora for different languages.

Scope and tasks: This special interest group is elaborating on suggestions for adapting the TEI guidelines to the representation of genres of computer-mediated communication (cmc). The focus of the group's work is on (but not limited to) tasks such as:

- modelling user contributions ("postings") to written cmc dialogues (which share features both with written discourse and with spoken utterances);
- modelling cmc document structures (*"cmc macrostructures"* – e.g., forum threads, wiki talk pages, chat logfiles, Twitter timelines etc.);
- annotating linguistic features within the posting element (*"cmc microstructures"* – elements such as emoticons, addressing terms, hashtags etc.);
- representing linked data and media objects connected with/embedded in cmc discourse;
- metadata schemata for the description of cmc resources;
- developing perspectives for the representation of discourse from multimodal cmc environments (in which written discourse, speech, video, animations and/or interaction in 3D worlds are used in combination).

The customized TEI schema and ODD document for CMC genres published in the *jTEI* (issue 3, 2012)[1] and first presented at the TEI members' meeting in Würzburg (2011) will serve as a starting point of the group's work.

Wiki space / schedule for getting started:

For the starting phase of their work, the group has set up a wiki with restricted access (wiki.itmc.tu-dortmund.de/cmc/). The results of the discussion within the group will be presented as feature requests on the TEI project pages on sourceforge.net – for a subsequent discussion of these suggestions with the members of

---

1 *Cf.* Michael Beißwenger, Maria Ermakova, Alexander Geyken, Lothar Lemnitzer, Angelika Storrer (2012): A TEI Schema for the Representation of Computer-mediated Communication. *Journal of the Text Encoding Initiative*, Issue 3 (November 2012): http://jtei.revues.org/476 (DOI: 10.4000/jtei.476).

The ODD document, schema documentation and encoding examples can be retrieved from http://www.empirikom.net/bin/view/Themen/CmcTEI

the TEI community. If the group is being accepted by the TEI Council as a new TEI SIG, the contents of the group wiki could be migrated to a wiki in the SIG section of the TEI website.

Contact person:

Michael Beißwenger, TU Dortmund University, Department for German Language and Literature, email: michael.beisswenger@tu-dortmund.de

Members of the core group:

The core group consists of representatives of corpus initiatives from four European countries which are currently building cmc corpora for several languages:

- Michael Beißwenger – TU Dortmund University (D)
- Thierry Chanier – Université Blaise Pascal, Clermont-Ferrand (F)
- Isabella Chiari – Università "La Sapienza", Rome (IT)
- Maria Ermakova – Berlin-Brandenburg of Sciences and the Humanities (D)
- Lothar Lemnitzer – Berlin-Brandenburg of Sciences and the Humanities (D)
- Angelika Storrer – TU Dortmund University (D)
- Henk van den Heuvel – Radboud University Nijmegen (NL)
- Maarten van Gompel – Radboud University Nijmegen (NL)

As soon as the core group will have put their first feature requests on sourceforget.net, the group shall be made open for anybody else who is interested to participate in the further discussion.

Activities and cooperations:

- The core group was formed as an outcome of the international workshop *"Building Corpora of Computer-Mediated Communication: Issues, Challenges, and Perspectives"* held Feb 13–15, 2013, at TU Dortmund University (funded by the Mercator Research Center Ruhr).[2]

- It is planned to submit a proposal for a panel at the TEI annual conference and members' meeting 2013 in Rome which gives an overview of challenges, existing suggestions and perspectives for the modelling of genres of computer-mediated communication in TEI.

- The SIG cooperates with the German scientific network "Empirical Research of Internet-Based Communication" (*Empirikom*[3]) funded by the German Research Foundation (DFG) as well as with the French SIG "Nouvelles formes des communication" (*Nouv-com*[4], part of the French national consortium *Corpusecrits*[5]) and with *CLARIN-NL*[6] (the Netherlands) and *CLARIN-D*[7] (Germany) which endorse TEI and other text annotation standards and support the exchange of mutual expertise for these standards.

---

2 For a documentation of the workshop *cf.* wiki.itmc.tu-dortmund.de/cmc/
3 http://www.empirikom.net/
4 https://groupes.renater.fr/wiki/corpus-ecrits-nouvcom/index
5 http://corpusecrits.corpus-ir.fr/
6 http://www.clarin.nl/
7 http://clarin-d.de/