Reykjavík (Iceland), 26th May 2014

Session 3

16:30-17:30

Chairperson: Thierry Declerck

16:30-17:00

Crowd-Sourced, Automatic Speech-Corpora Collection – Building the Romanian Anonymous Speech Corpus

Stefan Daniel Dumitrescu, Tiberiu Boroș, Radu Ion

Taking the example of other successful initiatives such as VoxForge, we applied the concept of crowd-sourcing to respond to a particular need: the lack of free-speech, time-aligned, multi-user corpora for the Romanian language. Such speech corpora represent a valuable asset for spoken language processing application because they offer the means to (1) train and test acoustic models and (2) develop and validate various methods and techniques that are intended to enhance today's ASR and TTS technologies.

17:00-17:30

Crowdsourcing for the Development of a Hierarchical Ontology in Modern Greek for Creative Advertising

Katia Kermanidis, Manolis Maragoudakis, Spyros Vosinakis

This paper describes the collaborative development of a hierarchical ontology in the domain of television advertisement. The language addressed is Modern Greek, one of the not widely spoken and not-richly-equipped-with-resources languages. The population of the ontology is achieved through collaborative crowdsourcing, i.e. players annotate ad video content through a multi-player videogame, implemented especially for this purpose. The provided annotations concern the

ad content, its production values, its impact, and they constitute the ontology terms and concepts. Dependencies, correlations, statistical information and knowledge governing the ontology terms and concepts are to be revealed through data mining and machine learning techniques. The extracted knowledge constitutes the core of a support tool, i.e. a semantic thesaurus, which will help ad designers in the brainstorming process of creating a new ad campaign. Unlike existing creativity support models, that are static and depend on expert knowledge, thereby hurting creativity, the proposed support tool is generic in nature (as it is based on a collaborative crowdsourcing-based semantic thesaurus), dynamic and minimally restricting the brainstorming process.