
Keynote Speech 1

Thursday, May 20th, 9:00

Chairperson: **Bente Maegaard**

Intelligence Resource Collection for Low-Density Languages

Jaime Carbonell

Whereas there are over 6,000 languages, less than 100 have ample resources that are easily collected. The next tranche of several hundred languages may have significant monolingual resources, but very little parallel text used to train corpus-based machine translation. Beyond the top 1,000 languages even monolingual text is rather scarce. This talk addresses methods to: 1) build language technologies methods, in particular MT methods, that require fewer scarce resources, and to 2) collect resources dynamically via active and proactive machine learning, so that the most crucial information is elicited when collection/annotation costs are high or information sources are otherwise limited. We focus on emerging new paradigms for MT based on richer linguistic priors, and consequent collection of human judgments primarily from bilinguals (e.g. translations, alignments, corrections, or heavier linguistic annotation) to supplement small-scale existing collections that may exist for the languages of interest.

Nom du document : Carbonell_abstract.doc
Répertoire : C:\Comm\Events\LREC\LREC 2010\Proceedings\BoA
Modèle : C:\Documents and Settings\Hélène\Application
Data\Microsoft\Modèles\Normal.dot
Titre :
Sujet :
Auteur : Irene Russo
Mots clés :
Commentaires :
Date de création : 21/04/2010 6:28
N° de révision : 3
Dernier enregistr. le : 26/04/2010 10:46
Dernier enregistrement par : Irene Russo
Temps total d'édition : 11 Minutes
Dernière impression sur : 30/04/2010 11:01
Tel qu'à la dernière impression
Nombre de pages : 1
Nombre de mots : 172 (approx.)
Nombre de caractères : 996 (approx.)