

Thematic Zonation and Text Visualisation

Brit Helle Aarskog

Dept. of Informatics (UiB)



Friday Nov 9, 2012, 10.15-11.15

VilVite Auditorium (TM51: AUD), Thormøhlensgate 51

Abstract:

Brit Helle Aarskog will present some central issues concerning thematic text zonation and procedures that determine named entities within these zones, thus providing data about the named entities' contextual scene.

A text zone constitutes a group of textual units with a certain number of shared features on lexical level, grammatical level, semantic level, and/or pragmatic level (statistics as dispersion values, mutual information score, keyness of keyness, etc.). A zonation procedure basically performs an advanced cluster analysis. The critical issue is to determine the threshold values for similarity scores. Thematic text zones are continuous, discontinuous, and overlapping with reference to the various levels of similarity.

When it comes to text visualisation, this is simply an open question which has not been

considered into detail. However, we have some basic ideas about for instance 'elevating' text zones with verbs mainly in the present tense, colour schemes for thematic zones (text spans with a high density of certain thematic indicators), and more wild ideas about how the user can select between several options for visualisation of textual patterns.

