



Hintergrund



<http://www.problogdesign.com>

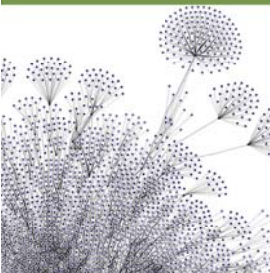
Aufgabe



www.onlinemarketingheringer.de/

References

Kontakt



<http://www.flickr.com/photos/maisonbisson/109211670/>

Diplomarbeit / Masterarbeit:

Computation of a web document complexity/readability score

Das Web enthält Dokumente mit unterschiedlichem Komplexitätsniveau. Es gibt Dokumente, die in einer leicht zu verstehenden Sprache geschrieben sind: kurze Sätze, wenige Fachbegriffe ohne Aufklärung, viele unterstützende Bilder. Solche Dokumente sind für Kinder und Erwachsene auch mit wenig Hintergrund im Thema der Dokumente geeignet. Andere Dokumente können in "fachlicher" Sprache geschrieben werden oder komplex strukturiert sein: viele spezifische Begriffe, komplexe Satzstruktur, keine Textgliederung. Selbst viele Erwachsene sind nicht in der Lage den Inhalt dieser Dokumente nachzuvollziehen. Die Unterscheidung der Dokumente ist wichtig für verschiedene Szenarios, z.B. die automatische Suche nach Webquellen geeignet für die Kinder oder die Unterstützung der Suche in einer Fremdsprache, wo es alleine aufgrund des mangelnden Sprachwissens kompliziert ist Dokumente zu verstehen.

Im Rahmen dieser Arbeit soll ein Maß vorgeschlagen werden, welches das Komplexitätsniveau von einem Webdokument bestimmt. Eine prototypische Implementierung in Java soll gemacht werden. Eine Evaluation des vorgeschlagenen Maßes soll gemacht werden um zu bewerten wie gut dieses für die Bewertung geeignet ist.

- ✓ N. Newbold, H. McLaughlin, L. Gillam. Rank by readability: Document weighting for information retrieval. Advances in Multidisciplinary Retrieval. Springer. 2010
- ✓ K. Collins-Thompson and J. Callan. A language modeling approach to predicting reading difficulty. Proceedings of HLT/NAACL 2004.
- ✓ S. Petersen and M. Ostendorf. A machine learning approach to reading level assessment. Computer, Speech and Language, vol. 23, no. 1, pp. 89-106, 2009

Wenn Sie Interesse oder Fragen haben, wenden Sie sich bitte an:

Tatiana Gossen
tatiana.gossen@ovgu.de
Büro: G29-118
Arbeitsgruppe Data and Knowledge Engineering