A Generalized Reuse Metrics (GRM) for Component Content Management Applications

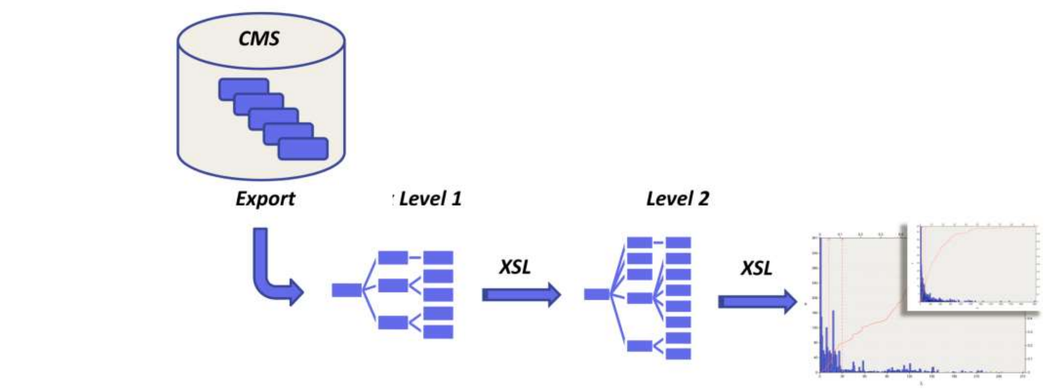
W. Ziegler

Karlsruhe University of Applied Sciences, Faculty of Information Management & Media

INTRODUCTION AND OBJECTIVES

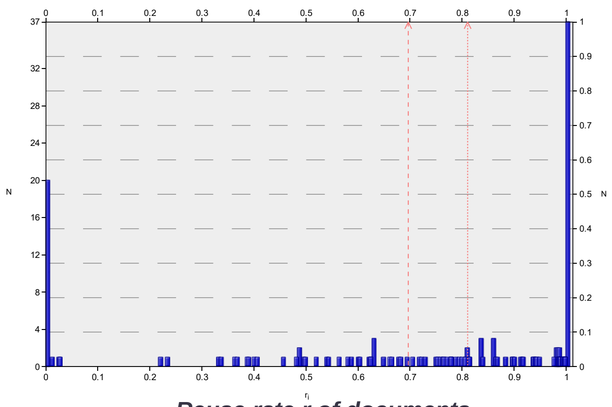
Component content management systems (CCMS) have become in certain global areas, for example in Central Europe, a standard toolset in technical communication (TC) [1]. They allow in a process-oriented and systematic way the creation and aggregation of content into document structures by reuse and subsequently publishing of deliverables in multiple media. However, in many cases, neither precise quality assurance processes are integrated in the workflow, nor standard mechanisms for gaining corresponding key performance indicators (KPI) describing CCMS process quality.

In order to improve this situation, corresponding KPI have been defined for the core mechanism of CCMS, i.e. for measuring the efficiency of content reuse within document structures [2]. From this, further basi and correlated quantities can be derived on a theoretical and practical basis. In contrast to other approaches [3], our methodology is independent of any specific implementation of XML information or content models or even of the data format. It is also independent of specific CCMS features. For this purpose, we solely rely on the generic concept of documents structures populated by referenced content objects building usually a tree structure of nested objects. This concept covers therefore the various types of standard or specialized CCMS implementations as well as standard frameworks like DITA-topics and corresponding maps [3]. By this, we developed a generalized reuse metrics (GRM). For practical applications, we use a generic XML exchange file format developed earlier. It has to be exported by the systems and carries CCMS information about content objects, document trees as XML data and corresponding metadata. Content itself is neither included in the export (level 1) nor in the analytics data (level 2) and KPI visualization.

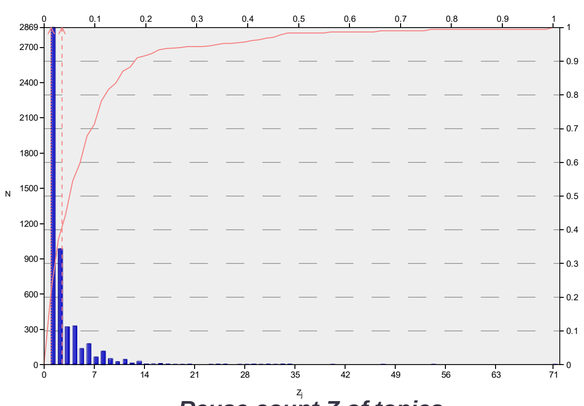


CONCEPT AND IMPLEMENTATION

From the export data file one can calculate all GRM quantities. There are two base quantities of a reuse metrics: The redundancy **r** of documents describing the relative amount of therein reused content, whereas the abundancy **Z** of content objects (usually topics) describes the number of reuses per object.
For a given system, these quantities can be depicted as distribution functions of the absolute numbers N(Z) and N(r) contain:



Reuse rate r of documents



Reuse count Z of topics

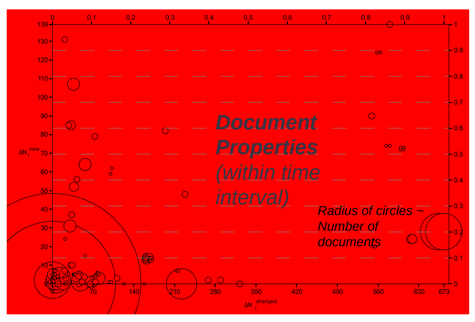
From this starting point and using metadata included in the export file, one can calculate for example the following static quantities for different object types:
Documents: number of contained words or topics, sharing factor (saving factor) as the weighted average of **Z** per topic contained in individual documents

Topics: number of revisions, number of contained words or fragments
Fragments: reuse count, fragment type
Media: reuse count, data size, number of revisions

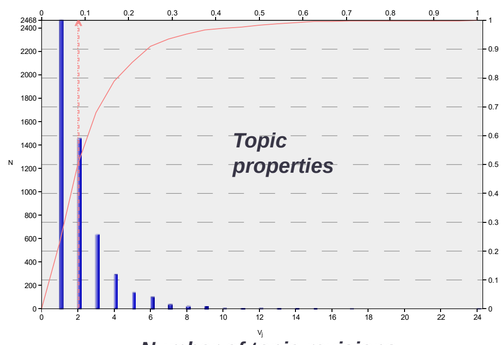
Dynamic Quantities
Investigating the time dependence of static quantities reveals the continious development of e.g. system use, content reuse and the evolution of content dynamics resp. document creation processes

Increase of document and topic numbers

Increase of created and reused content

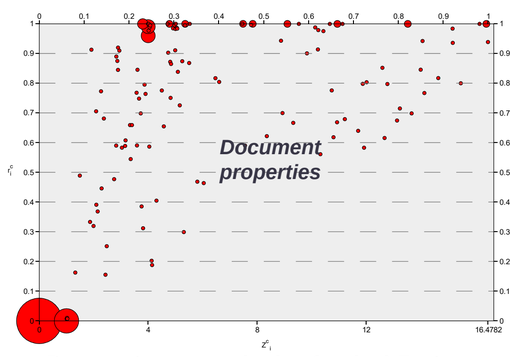


Number of changed topics vs. new topics

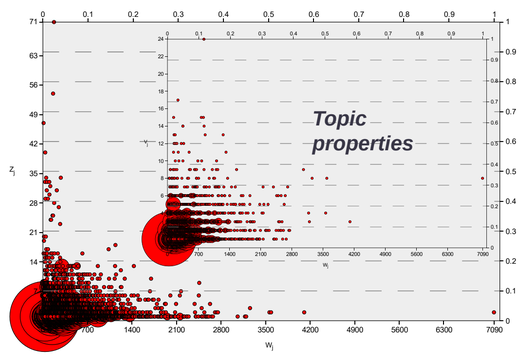


Number of topic revisions

Correlated Quantities
For detailed analysis, correlations of base quantities reveal dependencies and anomalies in object or system behavior.



Reuse rate versus sharing (saving) Factor



Topic size (words) correlated wiht reuse count Z and number of revisions (inset)

RESULTS AND OUTLOOK

- The GRM together with the export mechanism of raw CMS reuse data can be applied successfully to obtain relevant KPI for content management applications.
- CMS user companies obtain quantities of the GRM on different analytic levels: for management communication, for departmental workflow tracking and improvement and on a detailed authoring workgroup level among technical information creators
- For structured analysis and improvement of CMS processes (variant management, metadata use, versioning, reuse) data must be represented also as tabular data (e.g. spreadsheets) including corresponding metadata for filtering

REFERENCES

[1] D. Straub, W. Ziegler: “Effizientes Informationsmanagement durch komponentenbasierte Content-Management-Systeme (CCMS)”, tcworld (2019)

[2] W. Ziegler: “Metrische Untersuchung der Wiederverwendung im Content Management Statische Kennzahlen in der Technischen Redaktion” (2008)

[3] M. Lewis: “DITA Metrics 101”, Rockley Publishing (2012)

ACKNOWLEDGEMENTS

We acknowledge technical support by Claudia Oberle who implemented the numerical algorithms and performed the calculations. We acknowledge also the continuous cooperation with CCMS vendors Docufy, Expert Communication Systems, Fischer Information Technology and Schema and with all companies and system users providing their (anonymous) data for research.