



## Content Metrics for Products and Services Categorization Standards

Martin Hepp<sup>1,2</sup>, Joerg Leukel<sup>3</sup>, Volker Schmitz<sup>3</sup>

<sup>1</sup>Florida Gulf Coast University, Fort Myers, FL, USA

<sup>2</sup>Digital Enterprise Research Institute (DERI), Innsbruck, Austria

<sup>3</sup>University of Duisburg-Essen, Essen, Germany

2005 IEEE International Conference on e-Technology,  
e-Commerce and e-Service (EEE'05), March 29 – April 1, 2005

Dr. Martin Hepp, Research Cluster  
“Semantics in Business Information Systems”

<http://sebis.deri.org>

Digital Enterprise Research Institute (DERI), University of Innsbruck, Austria

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Full Paper / Citation Details

M. Hepp, J. Leukel, V. Schmitz: *Content Metrics for Products and Services Categorization Standards*, Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE-05), March 29 - April 1, 2005, Hong Kong, pp. 740-745.

Available at the IEEE Digital Library

<http://www.computer.org/publications/dlib/> or

IEEE Xplore

<http://ieeexplore.ieee.org/>

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Products and Services Categorization Standards (PSCS)

### What are PSCS?

**Classes, Taxonomies, and Properties, (...),**

reflecting domain consensus about how products or services can be grouped and described

### Importance and Core Applications

1. Any kind of mechanized content integration of product-related data (e.g. e-catalogs or spend analysis)

2. Parametric search

*=> A standard vocabulary for representing products and services in a form suitable for automated processing*

### Examples:

eCl@ss

<http://www.eclass.de>

UNSPSC

<http://www.unspsc.org>

eOTD

<http://www.eotd.org>

RosettaNet Technical Dictionary

<http://www.rosettanel.org>

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Metrics and Their Motivation

1. PSCS = **Architecture** + **Content**
2. ContentQuality(PSCS) = f(**Input**, **ProcessingPerformance**)
3. **Tagging** existing products and services data **is labor-intensive and costly**; you want do to it only once.
4. PSCS must cover the representational needs or your business – if there is not suitable category, you cannot tag your product (trivial, but important).
5. Users and standards bodies must be able to monitor the content quality (e.g. specificity, progress,...) of a given PSCS.
6. Simple metrics (e.g. number of classes) are of little value.

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Our Goal: Generic and Comprehensive Set of Metrics

Metrics that reflect the internal complexity (e.g. progress) and balance of any PSCS that meets common structural properties

- Size
- Growth and maintenance
- Hierarchical order and balanced content
- Property library
- Quality of class-specific property sets

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Size, Growth, and Maintenance

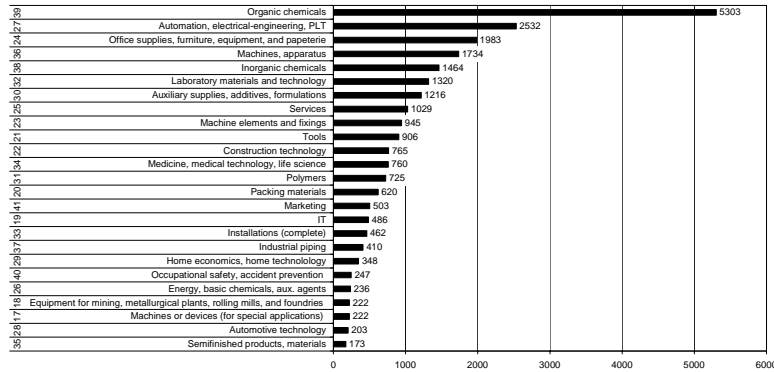
PSCS	Version	Total number of classes	New classes	Modified classes	New classes per 30 days	Modified classes per 30 days
eCI@ss	4.1	15315	n/a	n/a	n/a	n/a
eCI@ss	5.0	24814	13292	2418	865,0	157,4
eCI@ss	5.0SP1	24919	164	35	47,8	10,2
eCI@ss	5.1beta	25585	667	24918	131,6	4918,0
eCI@ss	5.1de	25658	84	0	74,1	0,0
eOTD	01-17-2003	58973	58973	n/a	n/a	n/a
eOTD	10-01-2003	58898	52	0	6,1	0,0
eOTD	11-01-2003	58901	5	0	4,8	0,0
eOTD	03-01-2004	58975	74	0	18,3	0,0
eOTD	06-01-2004	58970	5	0	1,6	0,0
eOTD	08-01-2004	58970	0	0	0,0	0,0

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



# Distribution of Entries along the Top-level Hierarchy

eCI@ss 5.0: Total Number of Nodes by Top-level Categories



Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



# Degree of Balance

	Release	% of classes in largest category	% of classes in 3 largest categories	Largest category / median of the category size
eCI@ss	4.1	23%	44%	814%
	5.0	21%	40%	731%
	5.0SP1	21%	40%	731%
	5.1beta	21%	39%	732%
	5.1de	21%	39%	732%
eOTD	10-01-2003	24%	40%	5255%
	11-01-2003	24%	40%	5254%
	03-01-2004	24%	40%	5255%
	06-01-2004	24%	40%	5255%
	08-01-2004	24%	40%	5255%
UNSPSC	6,0315	12%	30%	1128%
	6,0501	12%	29%	1134%
	6,0801	12%	30%	1134%
	6,1101	12%	30%	1108%
	7,0401	12%	30%	1107%
	7,0901	12%	30%	1107%

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Semantic Value and Weight

Semantic Weight for each attribute:

$$SW_i = \frac{1}{\text{Number of Attribute Lists Containing } A_i}$$

Semantic Value for each class:

$$SV_j = \sum SW_{A_i} \mid A_i \in S_j$$

Distribution properties!

Mean	0.17994013
Min	0.00019964
Max	87.88928283
Median	0.00025091
Mode	0.00025091

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Semantic Value

	Release	Semantic Value			
		Mean	Median	STD	Coefficient of Variation
eCI@ss	5.1de	4,74E-05	6,16E-07	2,48E-04	523%
eOTD	08.01.2004	1,70E-05	6,52E-09	7,32E-05	432%
RNTD	4.0	1,15E-03	6,41E-04	1,79E-03	155%

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Application

1. Corporations can use the metrics to evaluate the content quality of a specific descriptive language with regard to their product range or sourcing needs.

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



## Application

2. Standards bodies could use those metrics to
  - monitor the development of content quality,
  - assess the amount of resources necessary to eliminate the shortcomings,
  - rank content maintenance alternatives,
  - motivate industry groups to help improve currently weak segments of the standard.

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)



Thank you!

Martin Hepp

<http://sebis.deri.org>,  
<http://www.heppnetz.de>

How to contact the authors:

[martin.hepp@deri.org](mailto:martin.hepp@deri.org)

[joerg.leukel@uni-essen.de](mailto:joerg.leukel@uni-essen.de)

[volker.schmitz@uni-essen.de](mailto:volker.schmitz@uni-essen.de)

Martin Hepp, [martin.hepp@deri.org](mailto:martin.hepp@deri.org)