Document- and content-management: a thesaurus and onto-thesaurus approach

F.Vandamme*, M. Vandamme **, P. Kaczmarski*;

*BIKIT vzw, J. Plateaustr 22, 9000 Gent *LAE-UG, Blandijnberg 2, 9000 Gent

** VARTEC NV, Holstraat 19, 9000 Gent <u>Mike.Vandamme@vartec.be</u> <u>http://www.vartec.be/</u>

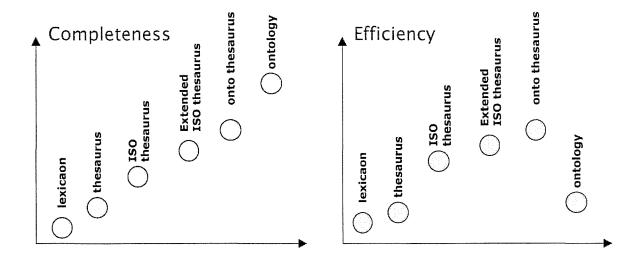
28 /04 /2005

Copyright (c) BIKIT 2005

Content

- Introduction and overview
 - DMS, CMS, ISO standards, thesaurus and onto-thesaurus
- Full-text versus thesaurus-based search
- Thesaurus and onto-thesaurus life cycle
- ◆ Creation of ISO thesauri : IKEM.ThesCreator[™]
- Maintenance of multilingual, multidomain ISO thesauri : IKEM.ThesEditor™
- Full-text, thesaurus,IKEM.Onto-thesaurus[™]
 and IKEM.QueryExpander[™]
- Live demonstration
- Conclusions

DMS/CMS knowledge-base options



Controlled vocabulary management

Developing, maintaining and managing controlled vocabularies (lexicons, standard ISO thesauri, ...) requires efforts depending on the vocabulary type

Problems, like

- Integration with document and content-managemet systems
- Completeness of the description
- Relevancy
- Consistency

IT solutions

- General groupware not well suited
- Dedicated software is a better option

Thesaurus standard: ISO 2788

ISO 2788: a mainstream standard for monolingual thesauri

The ISO 2788 standard defines an ISO thesaurus as a vocabulary of a controlled indexing language, organized so that the a priori relationships between concepts are made explicit.

ISO relation types

BT:

Broader term

NT:

Narrower term

RT:

Related term

USE:

Preferred term

UF:

Use For (non-preferred term)

SN:

Scope Note

ISO 2788: Thesaurus relations

hierarchic relations (broader/narrow (BT-NT) terms)

e.g., Asia --> broader term --> India

equivalent relations (use/use for (USE-UF) relations)

e.g., plastics -->use for --> polyethylene

related terms (associative (RT) relations)

e.g. vehicles--> rt --> wheels

• scope notes (examples, definitions (SN))

e.g. "audit client - person or organization commissioning the audit"

Thesaurus management: IKEM™

IKEM - Interactive Knowledge Management

A dedicated extensive, collaborative web-based software suite for

document- and content-management

- ISO 2788 compliant
- ISO 5964 compliant
- Optionally supports extensions beyond ISO

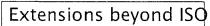
IKEM.TreeView
An example from
IKEM.ThesEditor 2005



Related Term (RT)

Use For (UF)

্র ৵ USE



Homonymy

Polyhierachy (multi-BT)



1

Full-text indexing: recall & precision

■ Precision

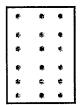
relevant retrieved
retrieved

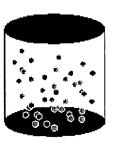
Recall

relevant retrieved

relevant in collection

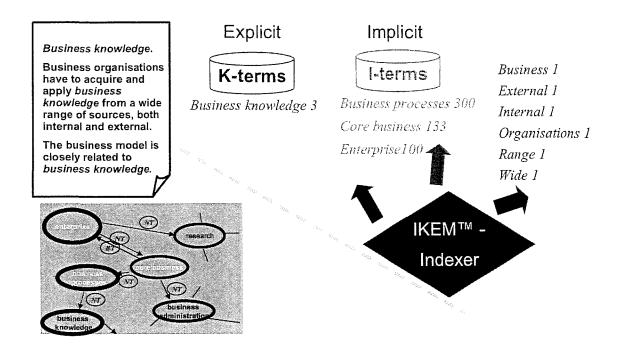
Retrieved Documents





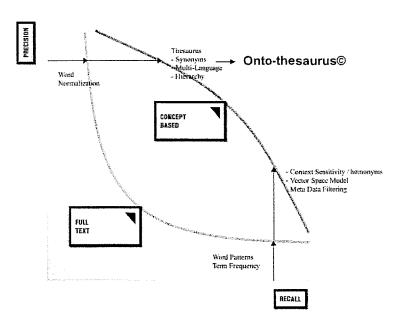
Collection

IKEM: Implicit knowledge

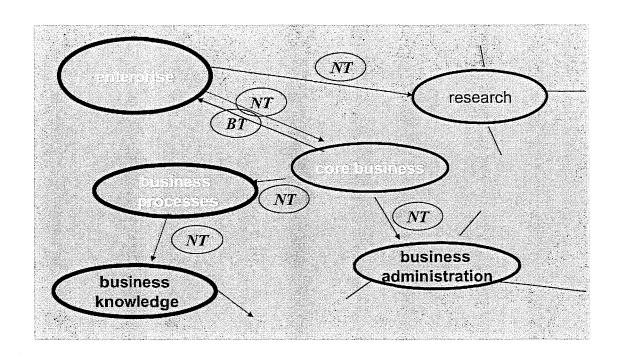


IKEM: Thesaurus & onto-thesaurus

 Thesauri and onto-thesauri offer an improvement of recall and precision



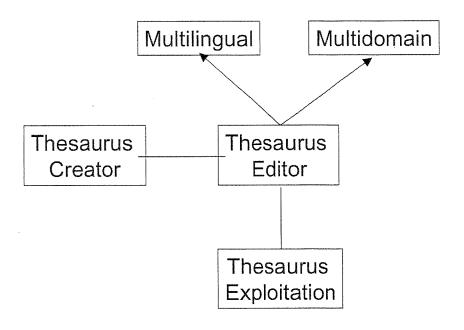
IKEM: implicit knowledge



15

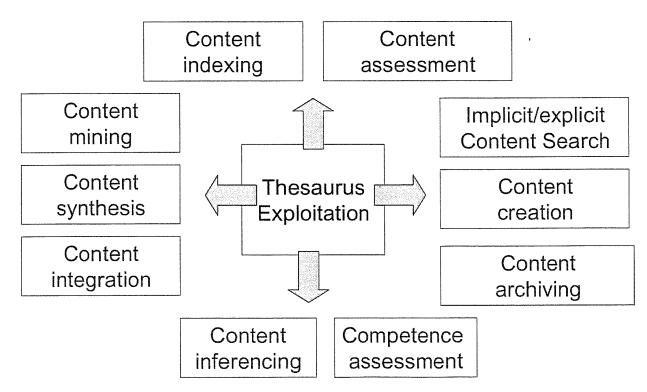
Thesaurus and onto-thesaurus life-cycle

Thesaurus life-cycle



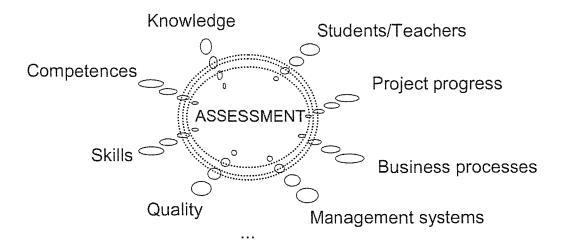
17

Thesaurus life-cycle



Thesaurus and onto-thesaurus based assessment

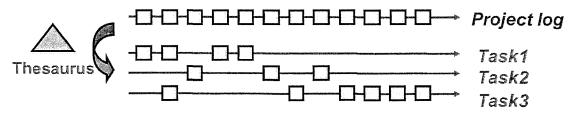
Assessment = any method used to better understand the present knowledge, skills or competencies of an individual, and the quality or characteristics of a process



Project assessment: message analysis

Project Views:

- · Task view
 - -dynamic overview of tasks as a function of time
 - -can be linked with project- and task specification
- Coordination view
 - -early warnings regarding possible delays in project realisation
- Cooperation view
 - -using actor profiles to analyze the cooperation patterns



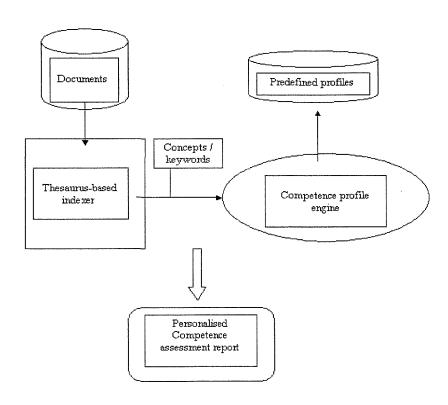
Competence assessment

Major issues

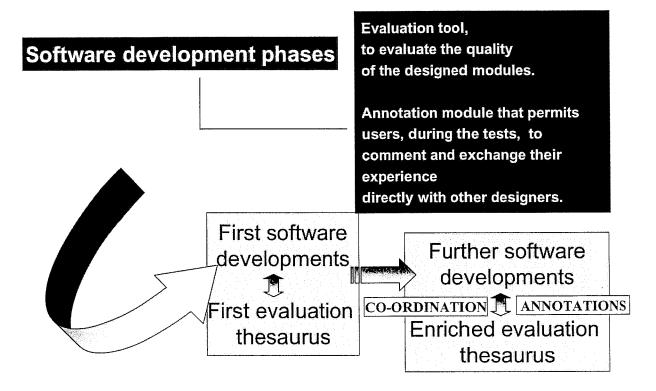
- Generate competence profiles based on e.g. the indexation of professional documents
- Compare with competence profiles in time & with predefined profiles
- Assess the evolution in time

Personalised competence assessment report

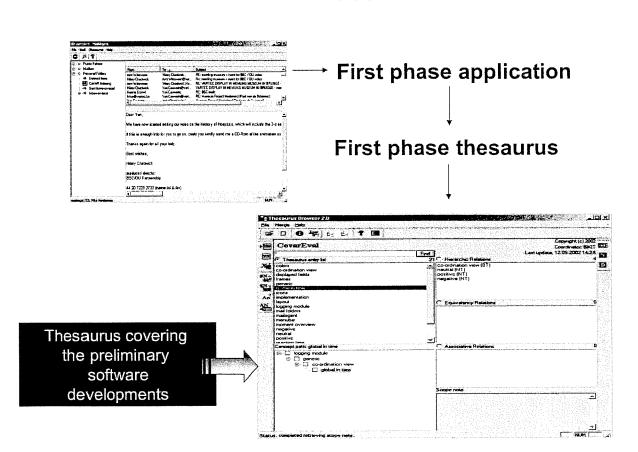
Competence assessment



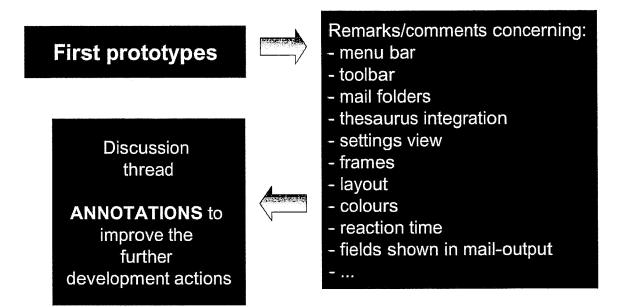
Software Assessment



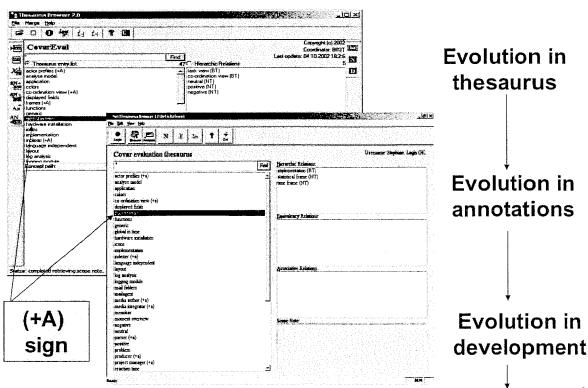
Software Assessment



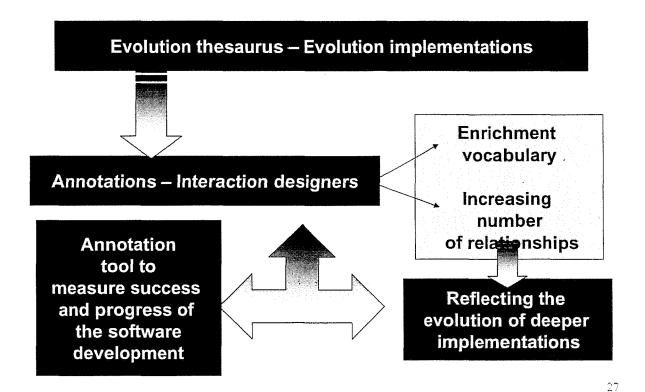
Software Assessment



Software Assessment

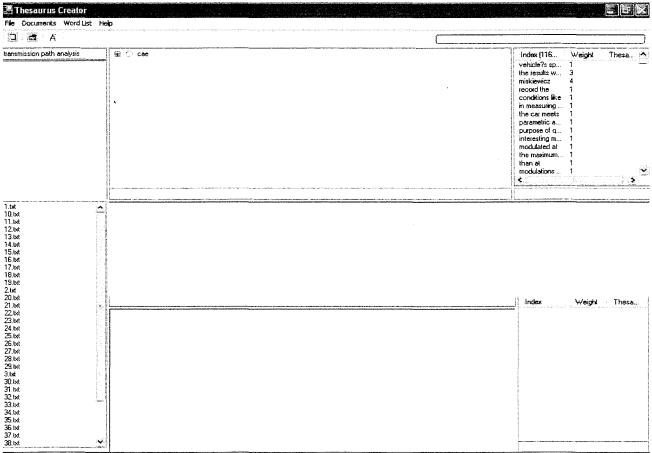


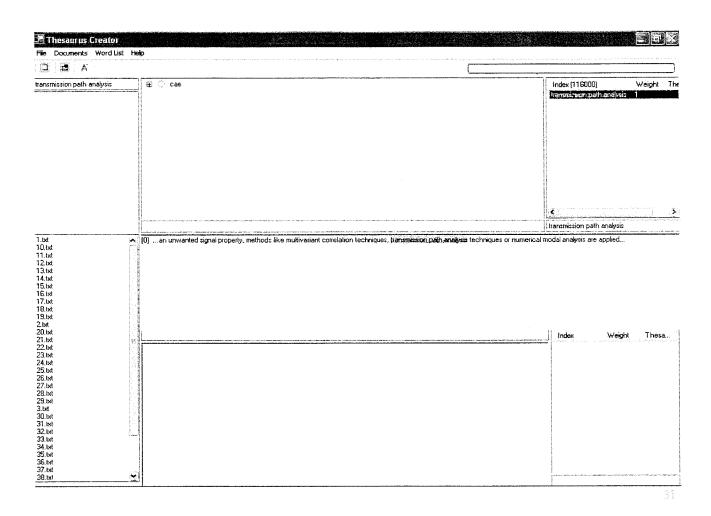
Software Assessment

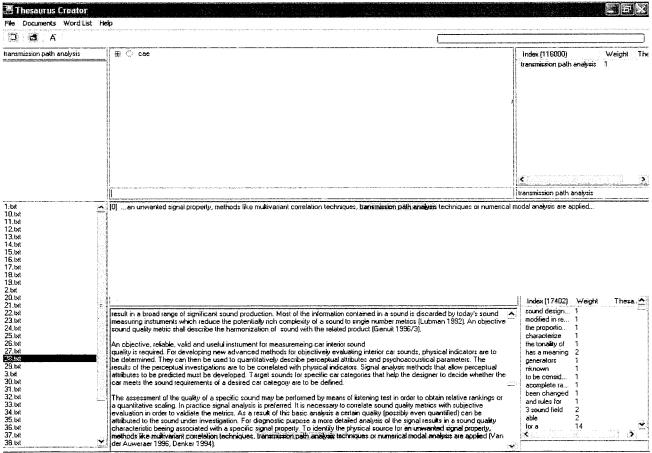


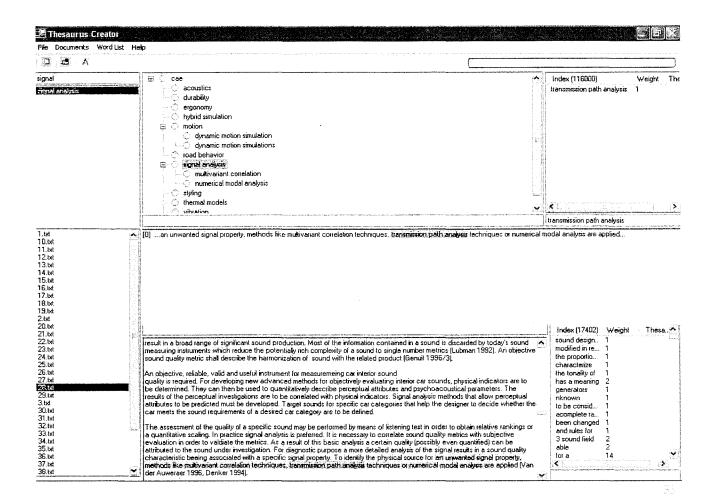
IKEM.ThesCreator ™ Creation of ISO thesauri

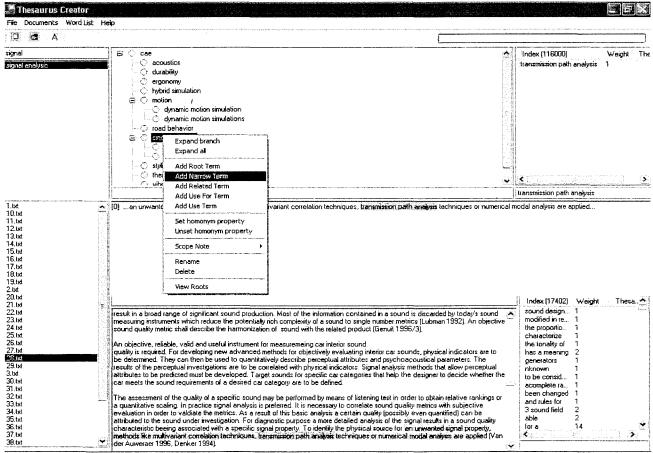
🇷 Thesaurus Creator	
File Documents Word List Help	
De Se A	
a cae coustics	Index (116 Weight Thesa ^ vehicle?s sp 1 the results w 3 miskiewicz 4 record the 1 conditions like 1 in measuring 1 the car meets 1 parametric a 1 purpose of q 1 interesting m 1 modulated at 1 the maximum 1 than at 1 modulations 1
1. bk 11. bk 11. bk 11. bk 12. bk 14. bk 15. bk 16. bk 17. bk 18. bk 19. bk 20. bk 20. bk 22. bk 22. bk 23. bk 24. bk 25. bk 26. bk 27. bk 28. bk 29. bk 30. bk 30. bk 30. bk 31. bk 32. bk 33. bk 33. bk 34. bk 35. bk 35. bk 35. bk 36. bk 37. bk 38. bk	Index Weight Thesa

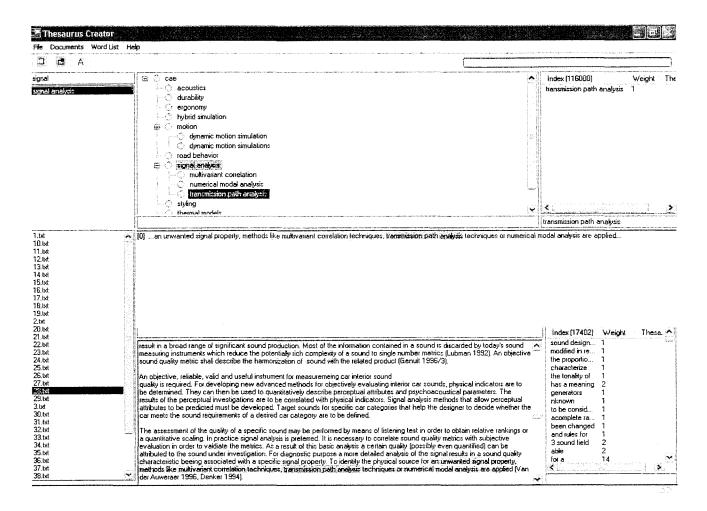


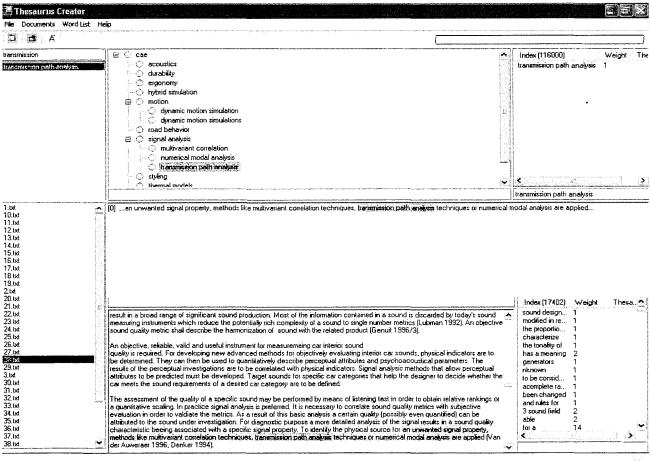












IKEM.ThesEditor[™]

Multilingual & multidomain thesaurus maintenance

37

IKEM.ThesEditor[™]2005

1 Thesaurus Viewing

- Tree browsing relations viewing
- List searching
- Concordance (single and multiple)
- Scope Notes viewing
- ISO script viewing

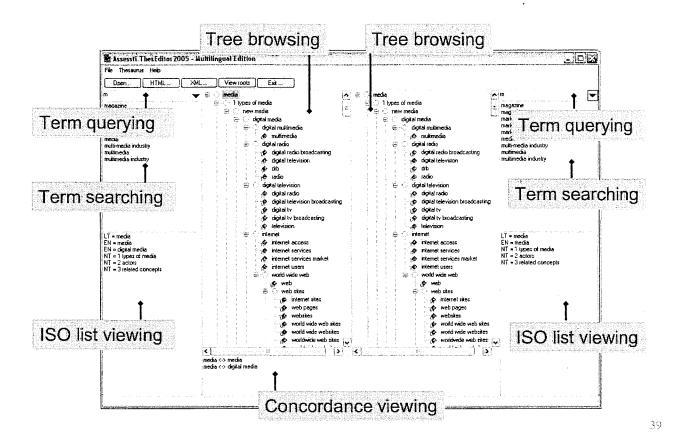
2 Thesaurus Editing

- Add / rename / delete related terms (BT/NT/RT/U/UF)
- Add / rename / delete Scope Note
- ISO script import

3 Thesaurus Export

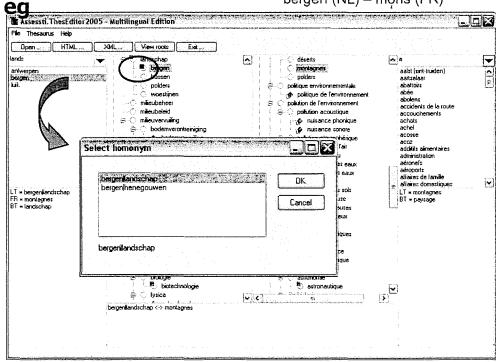
- Html
- XML

IKEM.ThesEditor[™]2005

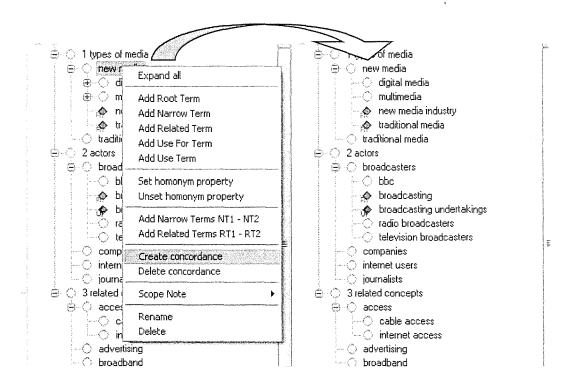


IKEM.ThesEditor[™]**2005**

Homonymy resolution bergen (NL) – montagnes (FR) bergen (NL) – mons (FR)



Concordance processing



4]

Onto-thesaurus application Searching a document library of train maintenance

