



The
Reuse
COMPANY



Innovation and Knowledge

The Reuse Company Trace > Reuse > Quality

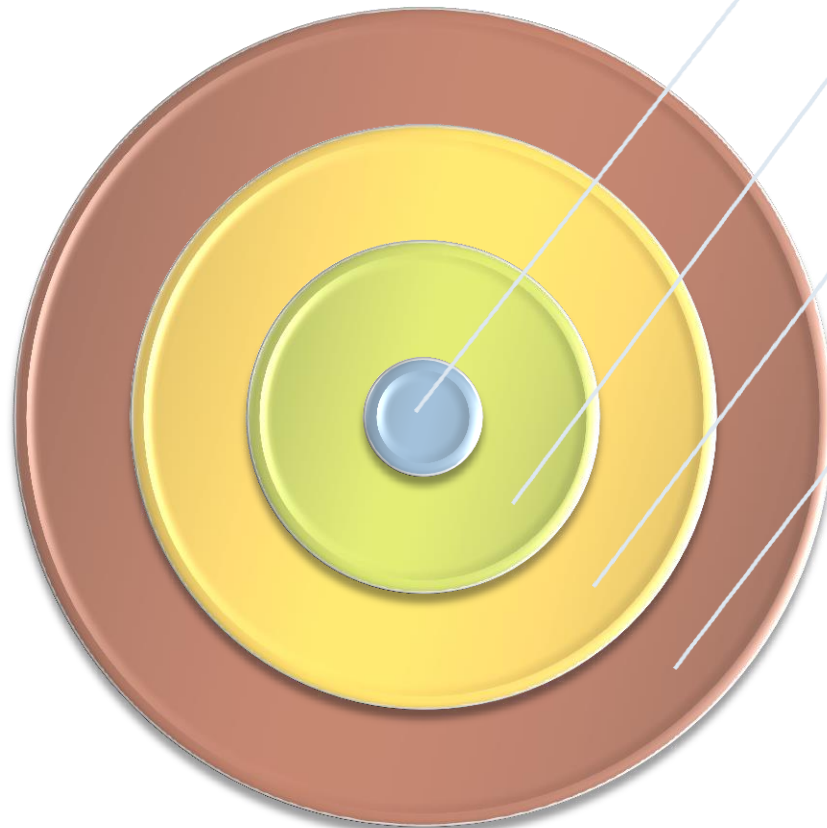


Ontology Construction :Vocabulary Management

www.kr.inf.uc3m.es
www.reusecompany.com

System Knowledge Base: Ontology

What is an ontology for TRC?



Controlled vocabulary: valid terms, forbidden terms... Optionally can include a Glossary (description for every term)

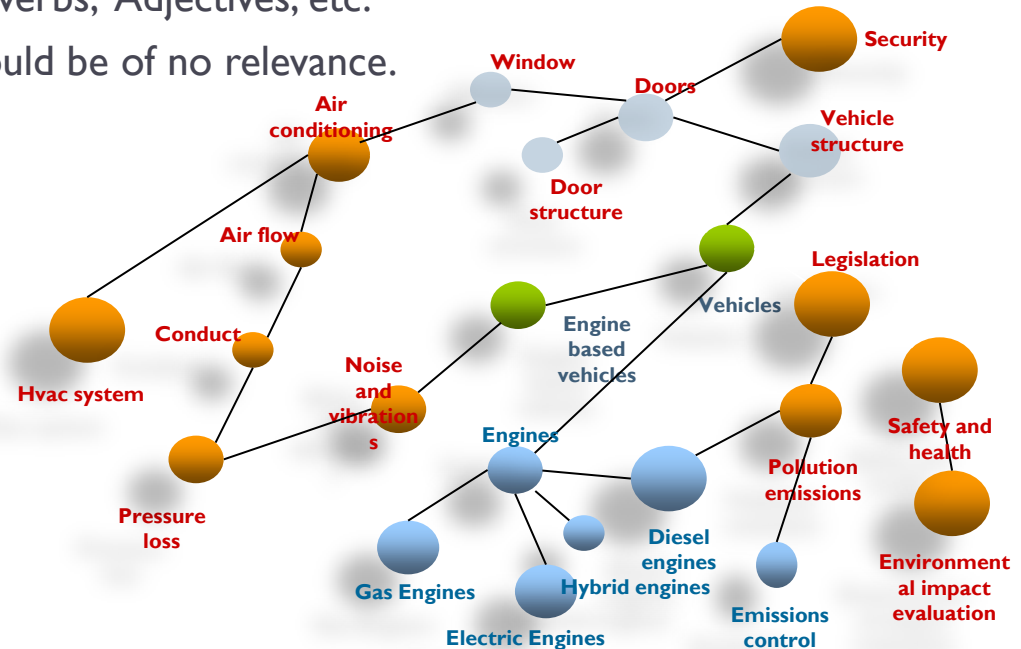
Thesaurus:
Relationships between terms: hierarchies, associations, synonyms...

Light Ontology:
syntactic and Semantic groupings for Terms and Actions (verbs). Domain terms and verbs

Patterns and Representation Schemas for Identifying (patterns) and representing (Schemas) the semantics of knowledge in Electronic Artifacts


Controlled Vocabulary

- Needed for standardizing and normalizing the terminology used in the custom application. The input information must/should match the controlled vocabulary.
- Using a glossary with different categories of terms, the ontology may store:
 - Business related Terms : those terms central to the business area to be treated
 - General Language Terms:
 - Syntactically relevant phrases: Adverbs, Adjectives, etc.
 - Invalid terms: those terms that could be of no relevance.



Controlled Vocabulary : Example for Requirements Authoring (RA)

UR044 :The Rad8 shall be able to identify hits at a minimum rate of 10 units per second

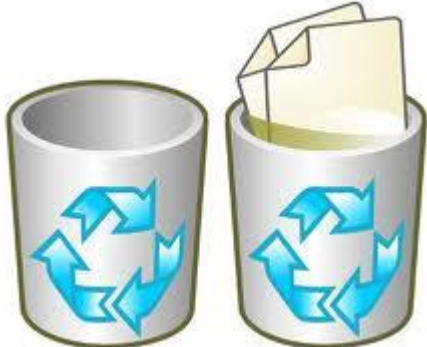


- shall
- identify
- hit
- minimum
- unit
-
- second


The

to

at



Rad8



What is knowledgeMANAGER

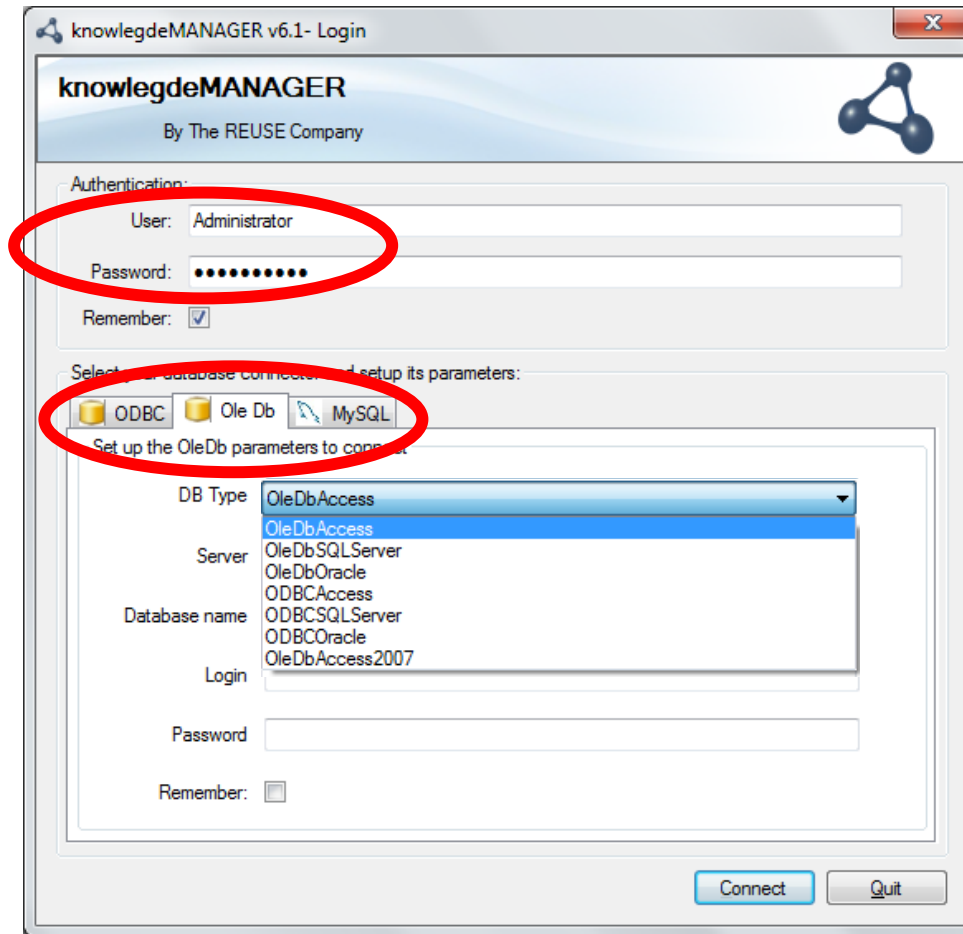


What is knowlegeMANAGER

- knowlegeMANAGER (kM) is a tool by TRC aimed to manage ontologies
- kM covers a bunch of different concepts:
 - Linguistic issues, vocabulary management, relationships, thesauri, patterns...
- kM (together with ontologies) represents the core of many TRC's projects
 - For RQA: managing vocabulary and relationships among concepts
 - For RAT: it allows to create boilerplates together with their formalization
 - For Semantic Search Engines and Reusable Repositories: it helps creating the semantics of the concepts

Connection screen and common interface

Connection screen



knowlegdeMANAGER v6.1- Login

knowlegdeMANAGER
By The REUSE Company

Authentication:

User: Administrator

Password:

Remember: ☒

Select a database connector and setup its parameters:

☒ ODBC ☐ Ole Db ☐ MySQL

Set up the OleDb parameters to connect:

DB Type: OleDbAccess

Server: OleDbSQLServer
OleDbOracle
ODBCAccess
ODBCSQLServer
ODBCOracle
OleDbAccess2007

Database name: OleDbAccess2007

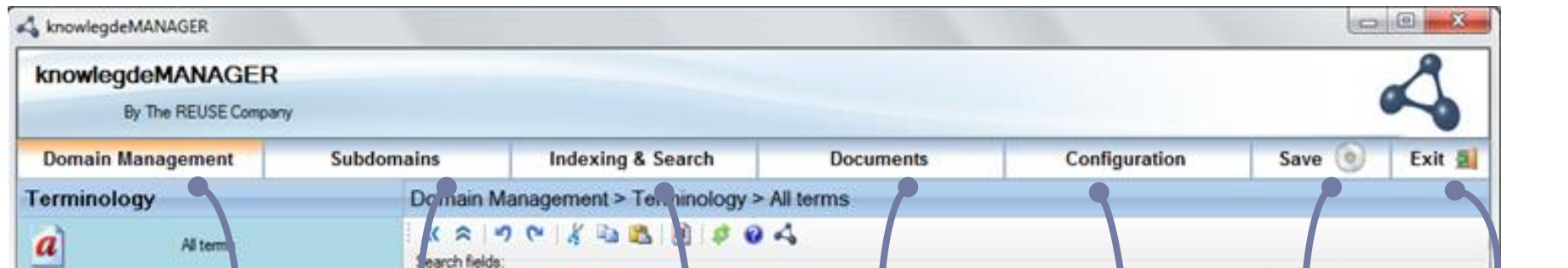
Login:

Password:

Remember: ☐

Connect Quit

Interface: Main menu



The screenshot shows the main menu of the knowlegdeMANAGER application. The menu items are: Domain Management, Subdomains, Indexing & Search, Documents, Configuration, Save, and Exit. The 'Terminology' sub-menu is expanded under 'Domain Management', showing 'All terms'. Arrows point from these menu items to their respective feature descriptions:

- Domain Management** (Terminology):
 - Vocabulary
 - Linguistic rules
 - Light ontology
 - Boilerplates
- Subdomains**:
 - Thesaurus
 - Suggestions
 - BP grouping
- Indexing & Search**:
 - Indexing tests:
 - Individual
 - Regression
- Documents**:
 - Asset management
- Configuration**:
 - Indexing configuration
 - Optimization processes:
 - Distances
 - Copying patterns
 - ...
- Save**:
 - kM storage:
 - A model in memory
 - Concurrent
- Exit**:
 - Exit

Filtering

knowlegdeMANAGER
By The REUSE Company

Domain Management | Subdomains | Indexing & Search | Documents | Configuration | Save | Exit

Terminology | Domain Management > Terminology > All terms

Special sentences
Add dictionaries
Term frequency analysis setup
Term frequency reports
Possible terminology inconsistencies

Identifier: ☐ Greater than: ☐ Lower than:

Term:

Tag:

Semantic:

Is in domain?: ☒ All terms in the terminology will be shown

Flags: ☐ Flag 1 ☐ Flag 2 ☐ Flag 3 ☐ Flag 4

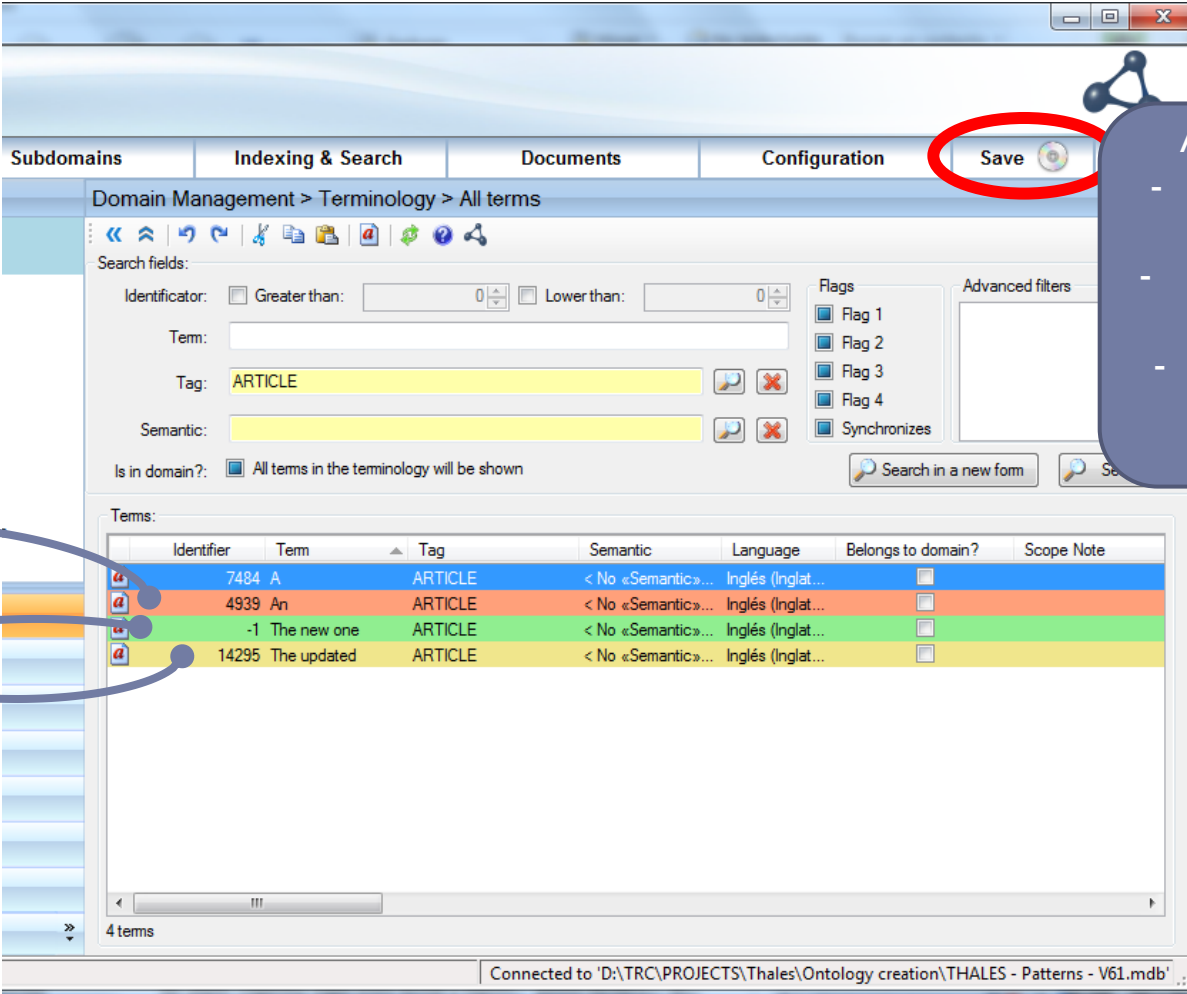
Advanced filters

Search in a new form Search

Identifier	Term	Tag	Semantic	Language	Belongs to domain?	Scope Note
10647	-	SUBSTRACTION	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10649	"	NOT_PUNCTUATIO...	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10650	#	NOT_PUNCTUATIO...	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10636	\$	NOT_PUNCTUATIO...	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10651	%	PERCENTAGE	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10652	&	NOT_PUNCTUATIO...	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
9521	(OPENING ROUND B...	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
9522)	CLOSING ROUND B...	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10664	*	NOT_PUNCTUATIO...	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10637	,	COMMA	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10642	.	SYMBOL	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10653	/	DIVISION	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
10654	:	SYMRO	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	

3.310 terms

Common interface: new, updated, removed items



The screenshot shows the 'Terminology' tab in the 'Domain Management' section. The 'Save' button is circled in red. The table below lists four terms with different background colors indicating their status: blue for 'Just added item', red for 'Removed item', green for 'Updated item', and yellow for 'Updated item'.

Identifier	Term	Tag	Semantic	Language	Belongs to domain?	Scope Note
7484	A	ARTICLE	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
4939	An	ARTICLE	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
-1	The new one	ARTICLE	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	
14295	The updated	ARTICLE	< No «Semantic»...	Inglés (Inglat...	<input type="checkbox"/>	

4 terms

Connected to 'D:\TRC\PROJECTS\Thales\Ontology creation\THALES - Patterns - V61.mdb'

After saving:

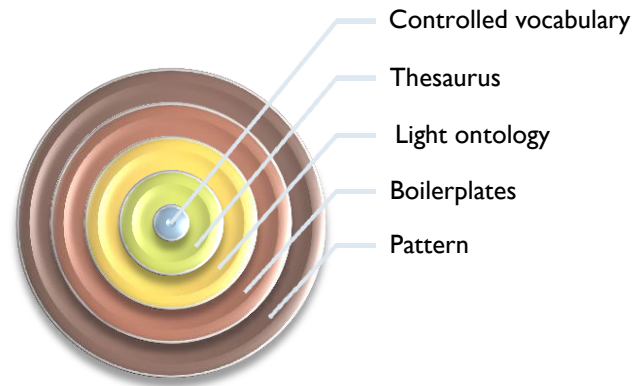
- Changes are stored
- Red rows will disappear
- Every other row will be white

Removed item

Just added item

Updated item

Vocabulary gathering

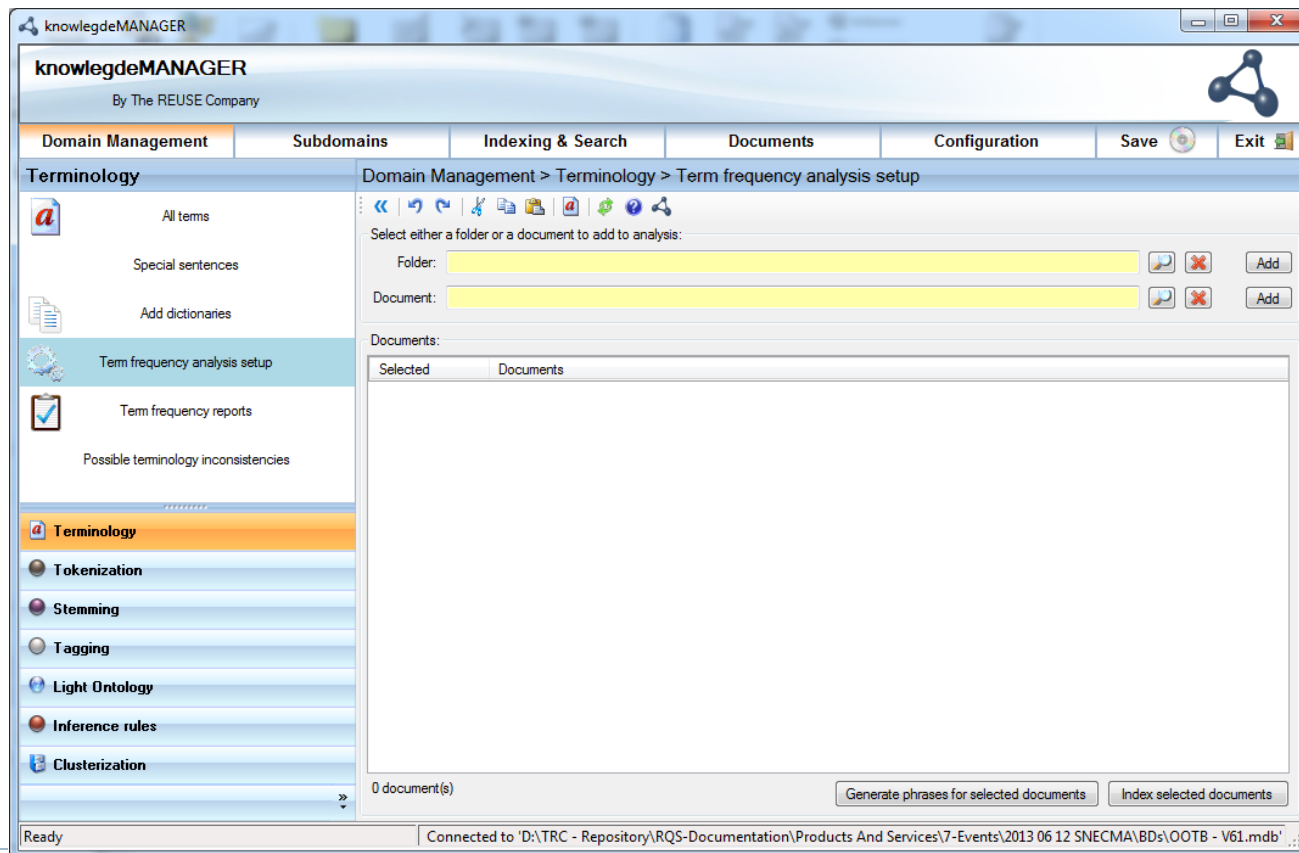


Gathering Vocabulary - To identify:

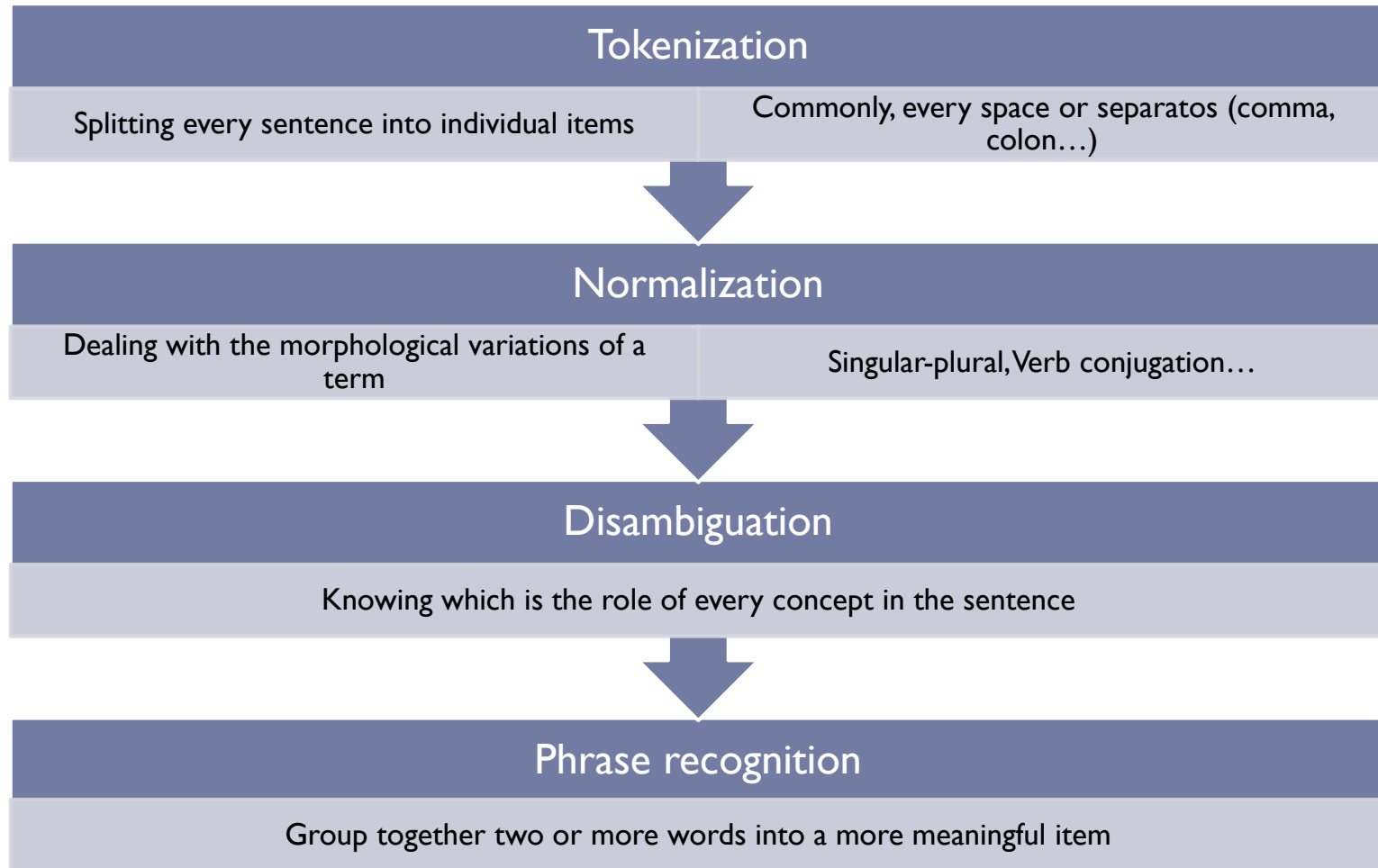
- Verbs Study and Identification
- Adjectives and Adverbs Study and Identification
- Acronyms identification
- Nouns identification
- Named Entities identification
- Simple terms identification
- Simple terms Validation
- Compound terms (phrases) Identification
- Verification and Validation of terms

Gathering Vocabulary – How to:

- Indexing Requirements or other documents
- An iterative process (do not be afraid of repeating the process)



Gathering Vocabulary : How it works



Vocabulary Gathering: How it works

All Radars shan't identify the following targeting enemy objectives:

Tokenization

[All] [Radars] [shall] [not] [identify] [the]
[following] [targeting] [enemy] [objectives] [:]

Normalization

[All] [Radar] [shall] [not] [identify] [the]
[following/follow] [targeting/Target] [enemy] [objective] [:]

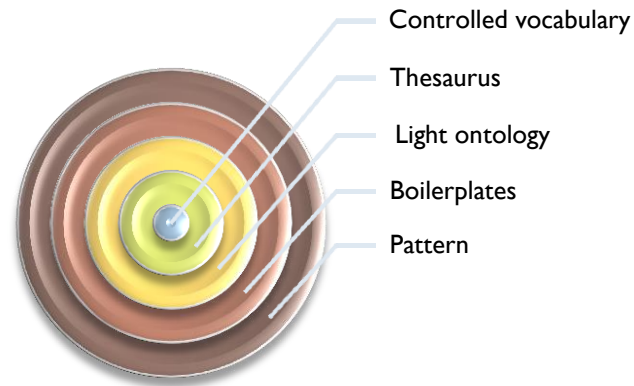
Disambiguation

[All] [Radar] [shall] [not] [identify] [the]
[following] [Target] [enemy] [objective] [:]

Phrase
Identification

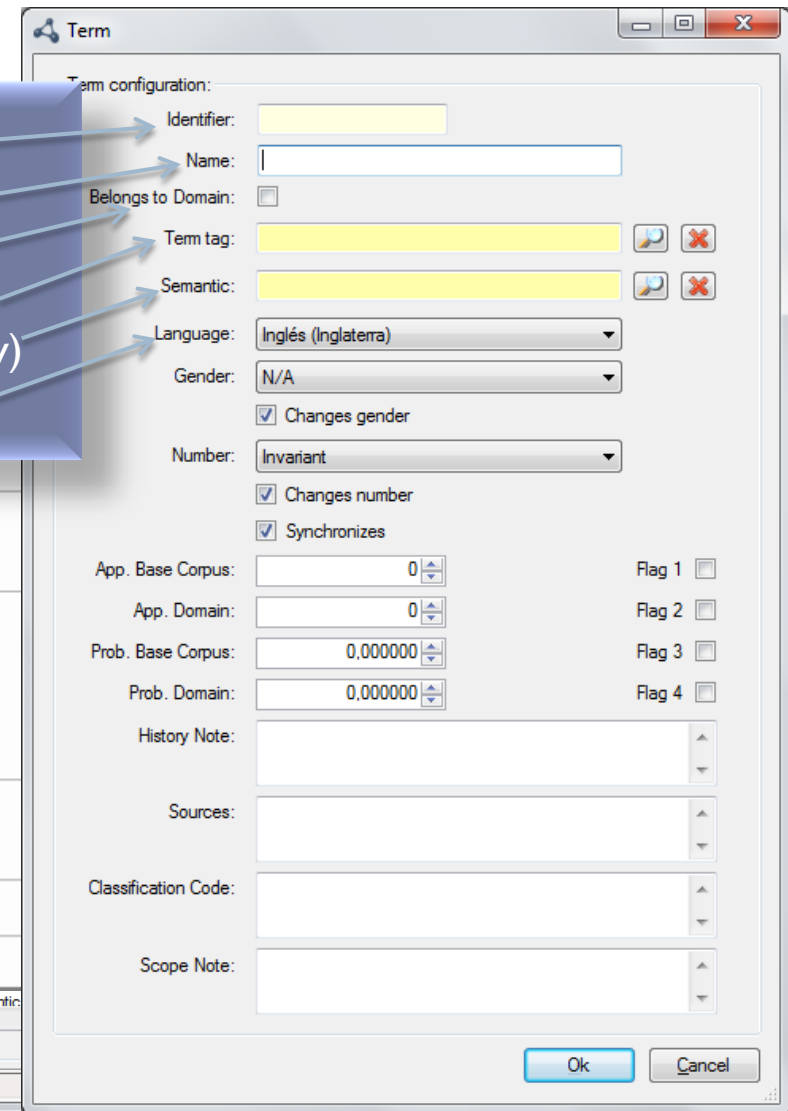
[All] [Radar] [shall] [not] [identify] [the]
[following] [Target] [enemy objective] [:]

Vocabulary management



Vocabulary management

Non-editable
 Mandatory
 See: Thesaurus
 Mandatory (one)
 Optional (1 or many)
 Mandatory







Term configuration:

Identifier:

Name:

Belongs to Domain: ☐

Term tag:  

Semantic:  

Language:


Gender:


☒ Changes gender


Number:


☒ Changes number

☒ Synchronizes

App. Base Corpus: 

App. Domain: 

Prob. Base Corpus: 

Prob. Domain: 

History Note:

Sources:

Classification Code:

Scope Note:

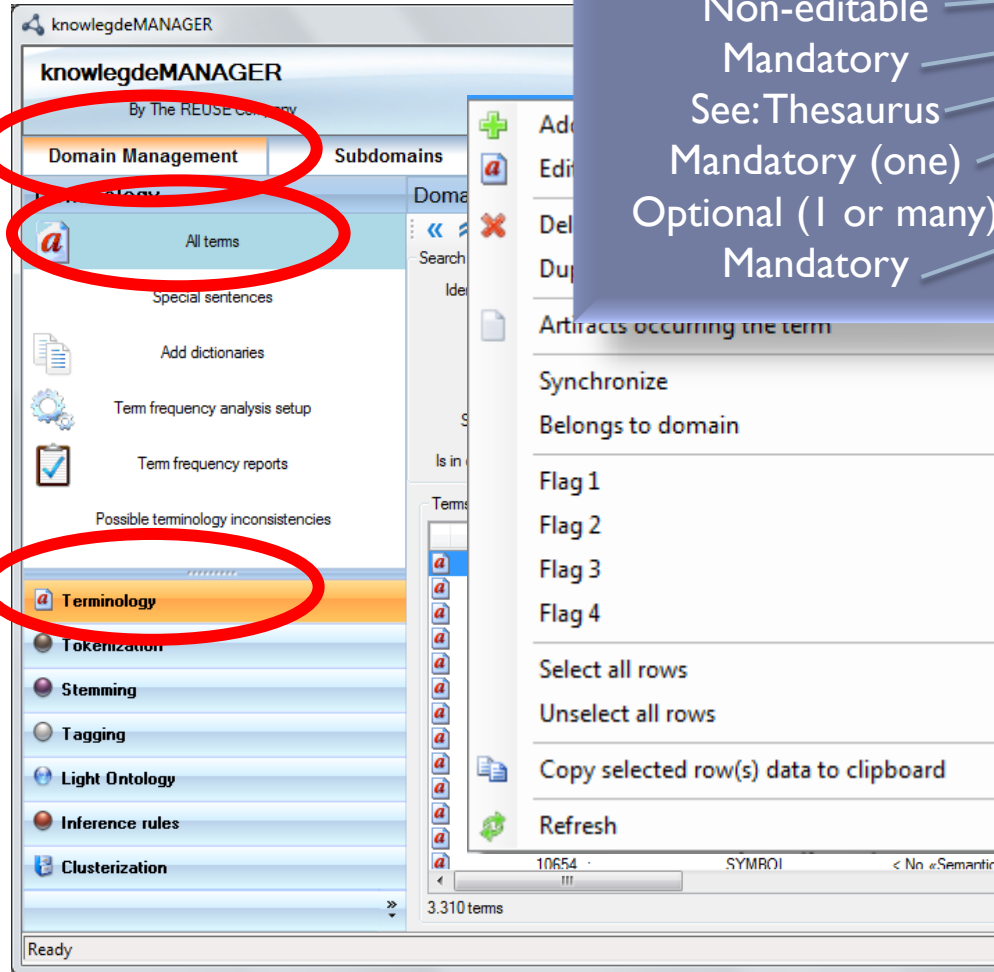
Flag 1 ☐

Flag 2 ☐

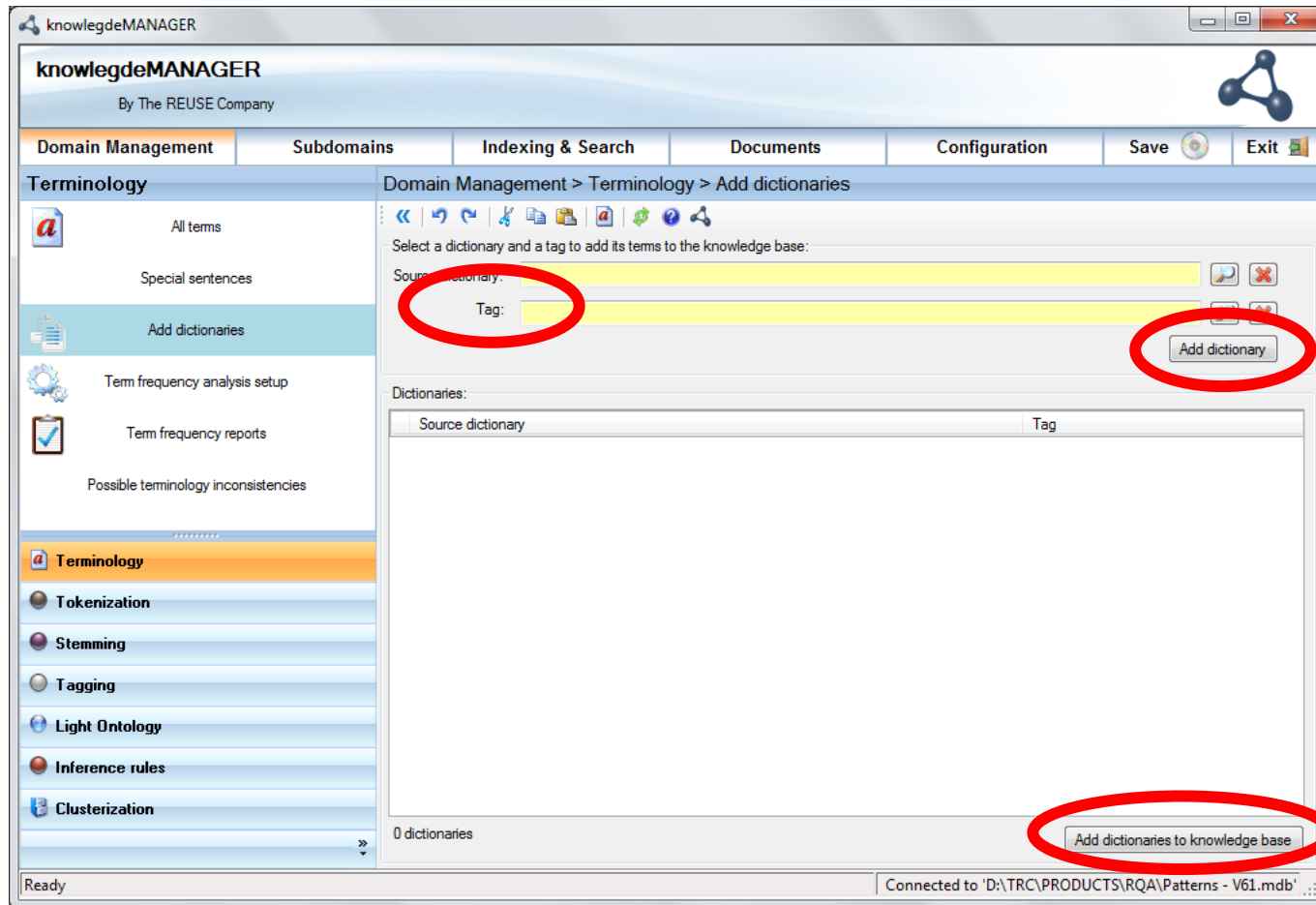
Flag 3 ☐

Flag 4 ☐

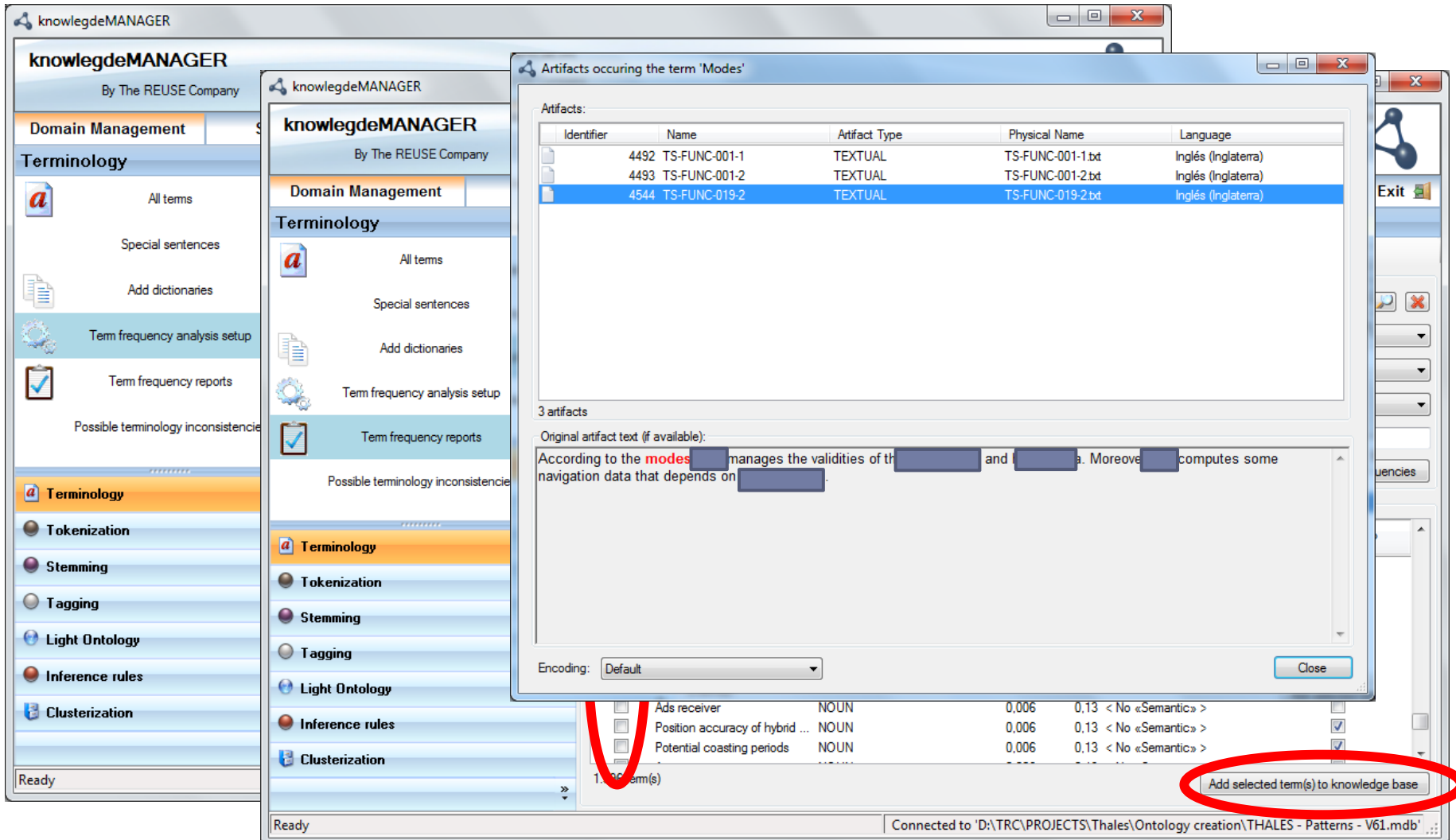
Ok Cancel



Vocaburaly: add dictionaries



Vocaburaly: freq analysis



knowlegdeMANAGER
By The REUSE Company

Domain Management

Terminology

- All terms
- Special sentences
- Add dictionaries
- Term frequency analysis setup
- Term frequency reports
- Possible terminology inconsistency

Tokenization

Stemming

Tagging

Light Ontology

Inference rules

Clusterization

Ready

Artifacts occurring the term 'Modes'

Artifacts:

Identifier	Name	Artifact Type	Physical Name	Language
4492	TS-FUNC-001-1	TEXTUAL	TS-FUNC-001-1.bt	Inglés (Inglaterra)
4493	TS-FUNC-001-2	TEXTUAL	TS-FUNC-001-2.bt	Inglés (Inglaterra)
4544	TS-FUNC-019-2	TEXTUAL	TS-FUNC-019-2.bt	Inglés (Inglaterra)

3 artifacts

Original artifact text (if available):

According to the **modes** manages the validities of the and . Moreover computes some navigation data that depends on .

Encoding: Default

Close

1. 100 term(s)

Add selected term(s) to knowledge base

Connected to 'D:\TRC\PROJECTS\Thales\Ontology creation\THALES - Patterns - V61.mdb'

Vocabulary: inconsistencies

knowledgeMANAGER

By The REUSE Company

Domain Management Subdomains Indexing & Search Documents Configuration Save Exit

Terminology

Domain Management > Terminology > Possible terminology inconsistencies

Search fields:

Term:

Term tag: **NOUN**

Is in domain?: ☒ All terms in the terminology will be shown

Is synonym?: ☐

Search in a new form Search

Terms whose name also normalizes with other terminology terms:

Term	Term Tag	Sema	Belongs to Domain?
MM	AC...	< N...	<input type="checkbox"/>
Mn	ME...	< N...	<input checked="" type="checkbox"/>
MN	AC...	< N...	<input type="checkbox"/>
Mode	NO...	< F...	<input checked="" type="checkbox"/>
Modes	NO...	< N...	<input type="checkbox"/>
Msec	ME...	< N...	<input checked="" type="checkbox"/>
MSEC	AC...	< N...	<input type="checkbox"/>
Needs	NO...	< N...	<input type="checkbox"/>
Nh	ME...	< N...	<input checked="" type="checkbox"/>
NH	AC...	< N...	<input type="checkbox"/>
Refueling system	NO...	< N...	<input type="checkbox"/>
Refuelling system	NO...	< N...	<input checked="" type="checkbox"/>
Rfs	NO...	< N...	<input type="checkbox"/>
Sae.ams1450	NO...	< N...	<input checked="" type="checkbox"/>

55 terms

Terms which also normalizes with 'Needs':

Term	Term Tag	Semant	Belongs to Domain?
Need	NOUN	< No ...	<input type="checkbox"/>

Delete selected term(s) from terminology

Merge term 'Category b' with term 'Category v' in the thesaurus and delete 'Category b' from terminology

Merge term 'Category b' with term 'Category v' in the thesaurus without deleting 'Category b' from terminology

Merge term 'Category v' with term 'Category b' in the thesaurus and delete 'Category v' from terminology

Merge term 'Category v' with term 'Category b' in the thesaurus without deleting 'Category v' from terminology

Select all rows

Unselect all rows

Copy selected row(s) data to clipboard

Refresh

1 terms

Ready

Connected to 'D:\TRC\PROJECTS\Thales\Ontology creation\THALES - Patterns - V61.mdb'

- This can happen when importing thesaurus
- Or normalization rules
- Or disambiguation rules

Tokenization

- Objective: to get tokens out of input text
- To correct possible misspelling problems
- To change input phrases
- To identify Named Entities:
 - Regular expressions

Tokenization Process

- 0 - Misspelling
 - Always first

- 1 - Sorted sequence of Regular Expressions
 - Order defined by the user

Misspelling Types

- ' PRE-PROCESSING
 - MS_PRE_Contiguous_Double_Chars = 1
 - MS_PRE_Word_Endings = 2
 - MS_PRE_Whole_Word = 3
- ' PROCESSING
 - MS_PROC_Contiguous_Double_Chars = 8
- ' IDENTIFICATION NUMBERS WRITTEN LETTERS
 - Identification_Words_To_Numbers = 9

Regular Expressions Types

- RE_SimpleReplace = 10
- RE_Groups_Replace = 11

- Remove_Symbols = 12
- Remove_Multiple_Spaces = 13

- Separate_Symbols = 14
- Separate_Symbols_From_Entities = 15

- RE_Named_Entities_Identification = 20
- RE_Named_Entities_Identification_With_Delimiters = 21

Tokenization

knowledgeMANAGER

By The REUSE Company

Domain Management Subdomains Indexing & Search Documents Configuration Save Exit

Tokenization Domain Management > Tokenization > Text Transformations

A=A' Text Transformations

Roles

Terminology

Tokenization

Stemming

Tagging

Light Ontology

Inference rules

Clusterization

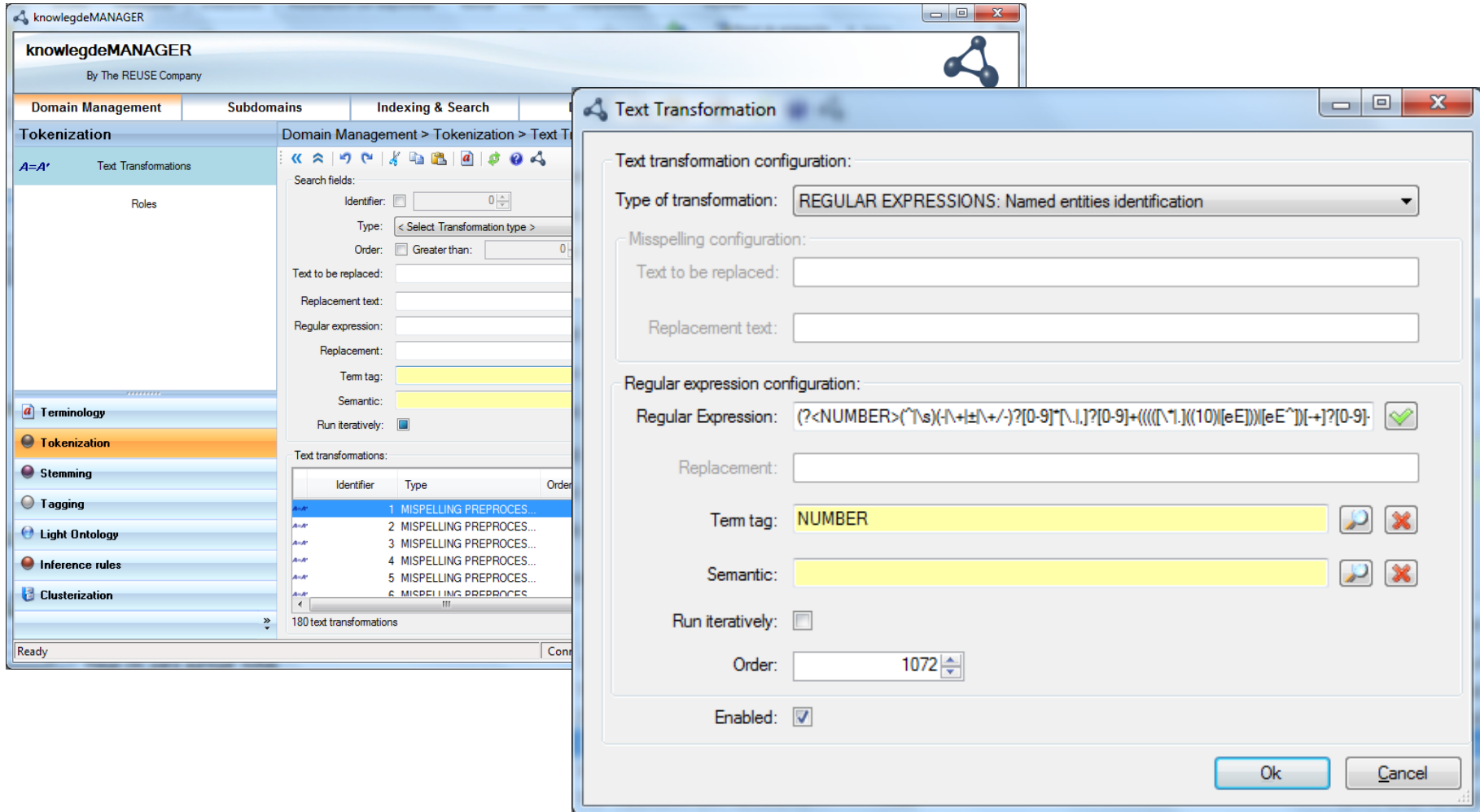
Text transformations:

Identifier	Type	Order	Text To Be Replaced	Replacement Text	Regular Expression
A=A'	1 MISPELLING PREPROCES...		and/or	andor	
A=A'	2 MISPELLING PREPROCES...		I'm	I am	
A=A'	3 MISPELLING PREPROCES...		you're	you are	
A=A'	4 MISPELLING PREPROCES...		he's	he is	
A=A'	5 MISPELLING PREPROCES...		she's	she is	
A=A'	6 MISPELLING PREPROCES...		we're	we are	
A=A'	7 MISPELLING PREPROCES...		they're	they are	
A=A'	8 MISPELLING PREPROCES...		it's	it is	
A=A'	9 MISPELLING PREPROCES...		won't	will not	
A=A'	10 MISPELLING PREPROCES...		shalln't	shall not	
A=A'	11 MISPELLING PREPROCES...		couldn't	could not	
A=A'	12 MISPELLING PREPROCES...		can't	can not	
A=A'	13 MISPELLING PREPROCES...		aren't	are not	
A=A'	14 MISPELLING PREPROCES...		didn't	did not	
A=A'	15 MISPELLING PREPROCES...		don't	do not	
A=A'	16 MISPELLING PREPROCES...		hadn't	had not	
A=A'	17 MISPELLING PREPROCES...		hasn't	has not	
A=A'	18 MISPELLING PREPROCES...		haven't	have not	
A=A'	19 MISPELLING PREPROCES...		isn't	is not	
A=A'	20 MISPELLING PREPROCES...		mightn't	might not	
A=A'	21 MISPELLING PREPROCES...		mustn't	must not	
A=A'	22 MISPELLING PREPROCES...		needn't	need not	






179 text transformations

Ready Connected to 'D:\TRC - Repository\RQS-Docummentation\Products And Services\7-Events\2013 06 12 SNECMA\BDs\OOTB - V61.mdb'

Tokenization



The screenshot shows the **knowlegdeMANAGER** interface with the **Text Transformation** dialog box open. The main window displays the **Tokenization** section under **Domain Management**. The **Text Transformation** dialog box is configured as follows:

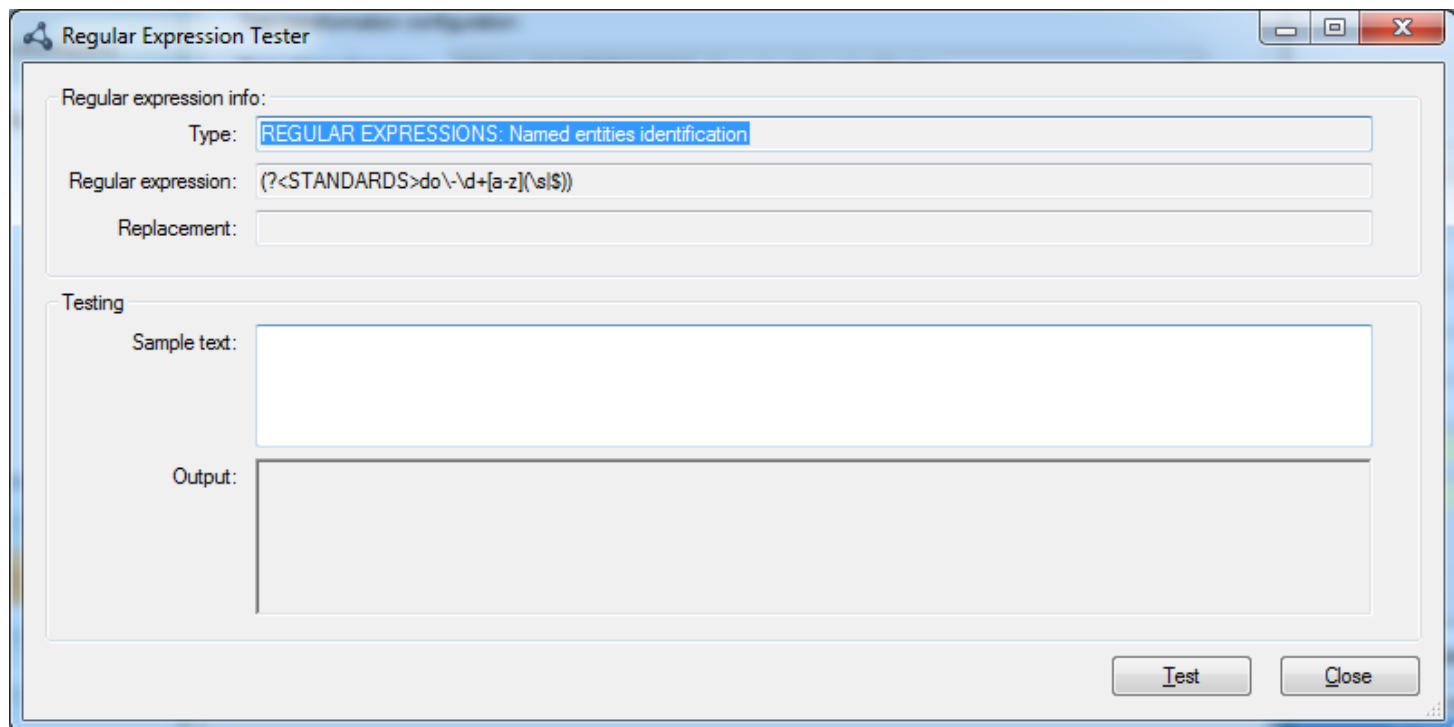
- Text transformation configuration:**
 - Type of transformation: **REGULAR EXPRESSIONS: Named entities identification**
- Misspelling configuration:**
 - Text to be replaced: (empty)
 - Replacement text: (empty)
- Regular expression configuration:**
 - Regular Expression: `(?<NUMBER>(^\\s)(-\\+|\\+/-)?[0-9]*[\\.]?[0-9]+(((\\^1)(10)[eE]))[eE^][+]?[0-9])` 
 - Replacement: (empty)
 - Term tag: **NUMBER**  
 - Semantic: (empty)  
 - Run iteratively: ☐
 - Order: **1072**
 - Enabled: ☒

The background window shows the **Text Transformations** list with 180 transformations. The first six are listed:

Identifier	Type	Order
1	MISPELLING PREPROCES...	
2	MISPELLING PREPROCES...	
3	MISPELLING PREPROCES...	
4	MISPELLING PREPROCES...	
5	MISPELLING PREPROCES...	
6	MISPELLING PREPROCES...	

Tokenization : Testing Regular Expressions

- NOT valid regular expressions presuming Upper case input
- The indexer transforms everything to LOWER



A screenshot of a software window titled "Regular Expression Tester". The window has a standard Windows-style title bar with minimize, maximize, and close buttons. Inside, there are two main sections: "Regular expression info:" and "Testing".

The "Regular expression info:" section contains three input fields:

- Type:** A text box containing "REGULAR EXPRESSIONS: Named entities identification".
- Regular expression:** A text box containing "(?<STANDARDS>do\\-\\d+[a-z](\\s|\$))".
- Replacement:** An empty text box.

The "Testing" section contains two input fields:

- Sample text:** An empty text box.
- Output:** A larger empty text box.

At the bottom right of the window, there are two buttons: "Test" and "Close".

Tokenization Problems

- The tokenization process can produce strange results when a domain changes
 - Ex: the following RegEX: `(?<1>[a-z]+)(?<2>[\-/])(?<3>[0-9a-z]+)`
 - works well in English language to separate units and words by /, like in Km/h, or Input/Output , but if it is used in French, the Arrêt/Démarrage, would produce a strange result Arrê t / D émarrage
 - Solution:
 - `(?<1>[a-záéíóúàèìòùâêîôûäëïöü]+)(?<2>[\-/])(?<3>[0-9a-záéíóúàèìòùâêîôûäëïöü]+)`
- The tokenization process can produce strange results, depending on the order of application of the regular expressions

External tools:

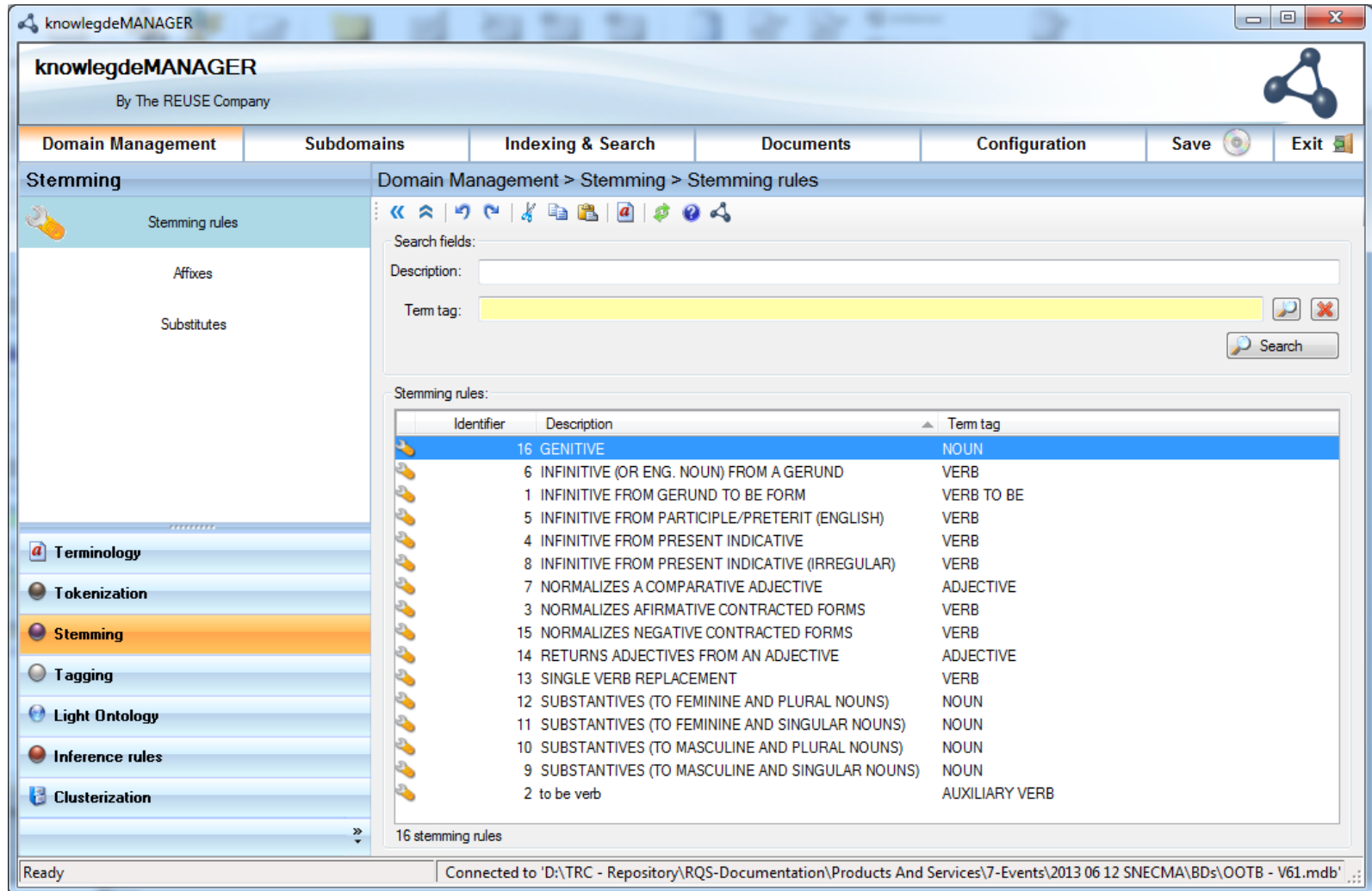


External tools:



The screenshot shows the RegExLib.com website in a web browser. The browser's address bar displays <http://www.regexlib.com/>. The website has a green header with the logo "RegExLib.com Regular Expression Library" and a search bar containing the regex `^[a-zA-Z]+$`. A navigation menu includes links for Home, Search, Regex Tester, Browse Expressions, Add Regex, and Login. Below the menu is a Google search bar with the text "Tus clientes potenciales" and a "Please support RegExLib Sponsors" message. The main content area is divided into several sections: "Subscribe" with a "Recent Expressions" link showing 194 readers by Feedburner; "Site Links" with a list of links including "Regex Cheat Sheet", "Search", "Regex Tester", "Browse Expressions", "Add Regex", "Manage My Expressions", "Contributors", "Regex Resources", "Web Services", "Advertise", "Contact Us", "Register", "Recent Expressions", and "Recent Comments"; "Find Expressions" with a search bar and a "Search" button; "Regex Resources" featuring "The Regulator" tool, described as a comprehensive .NET tool for testing Regular Expressions, with a link to "View all Regular Expression resources..."; and "Latest News". A "Community" section at the bottom left links to "Regex Forums" and "Regex Blogs".

Normalization / Stemming



knowlegdeMANAGER
By The REUSE Company

Domain Management | Subdomains | Indexing & Search | Documents | Configuration | Save | Exit

Stemming

Stemming rules

Affixes

Substitutes

Terminology

Tokenization

Stemming

Tagging

Light Ontology

Inference rules

Clusterization

Domain Management > Stemming > Stemming rules

Search fields:

Description:

Term tag:

Search

Stemming rules:

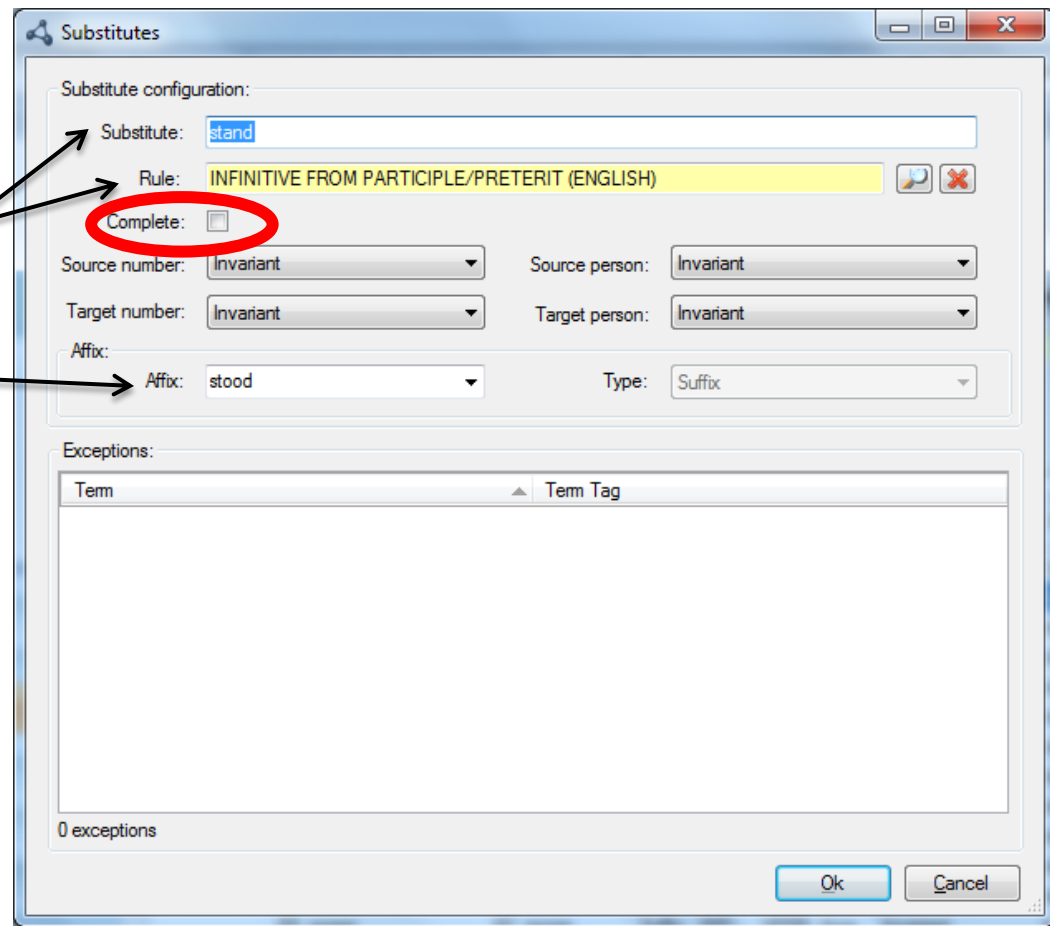
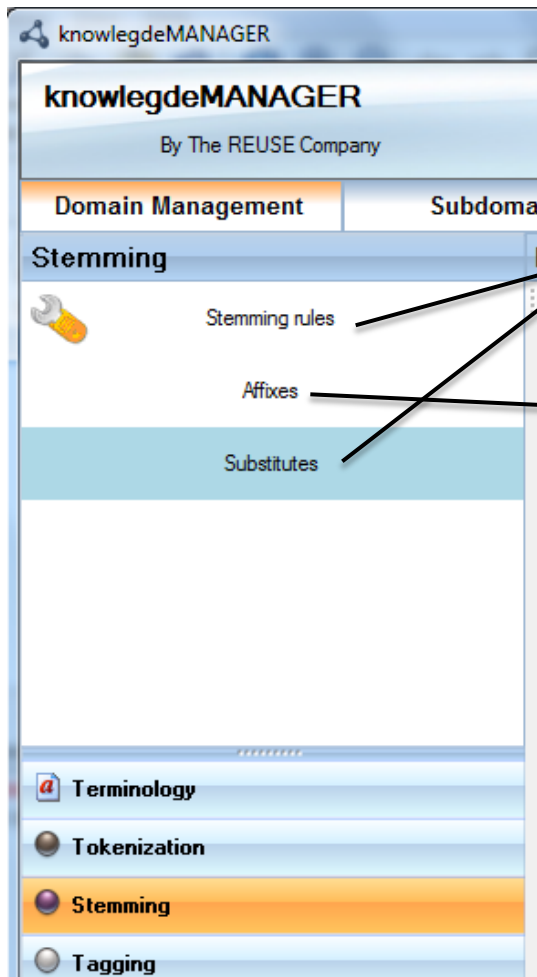
Identifier	Description	Term tag
16	GENITIVE	NOUN
6	INFINITIVE (OR ENG. NOUN) FROM A GERUND	VERB
1	INFINITIVE FROM GERUND TO BE FORM	VERB TO BE
5	INFINITIVE FROM PARTICIPLE/PRETERIT (ENGLISH)	VERB
4	INFINITIVE FROM PRESENT INDICATIVE	VERB
8	INFINITIVE FROM PRESENT INDICATIVE (IRREGULAR)	VERB
7	NORMALIZES A COMPARATIVE ADJECTIVE	ADJECTIVE
3	NORMALIZES AFIRMATIVE CONTRACTED FORMS	VERB
15	NORMALIZES NEGATIVE CONTRACTED FORMS	VERB
14	RETURNS ADJECTIVES FROM AN ADJECTIVE	ADJECTIVE
13	SINGLE VERB REPLACEMENT	VERB
12	SUBSTANTIVES (TO FEMININE AND PLURAL NOUNS)	NOUN
11	SUBSTANTIVES (TO FEMININE AND SINGULAR NOUNS)	NOUN
10	SUBSTANTIVES (TO MASCULINE AND PLURAL NOUNS)	NOUN
9	SUBSTANTIVES (TO MASCULINE AND SINGULAR NOUNS)	NOUN
2	to be verb	AUXILIARY VERB

16 stemming rules

Ready

Connected to 'D:\TRC - Repository\RQS-Documantation\Products And Services\7-Events\2013 06 12 SNECMA\BDs\OOTB - V61.mdb'...

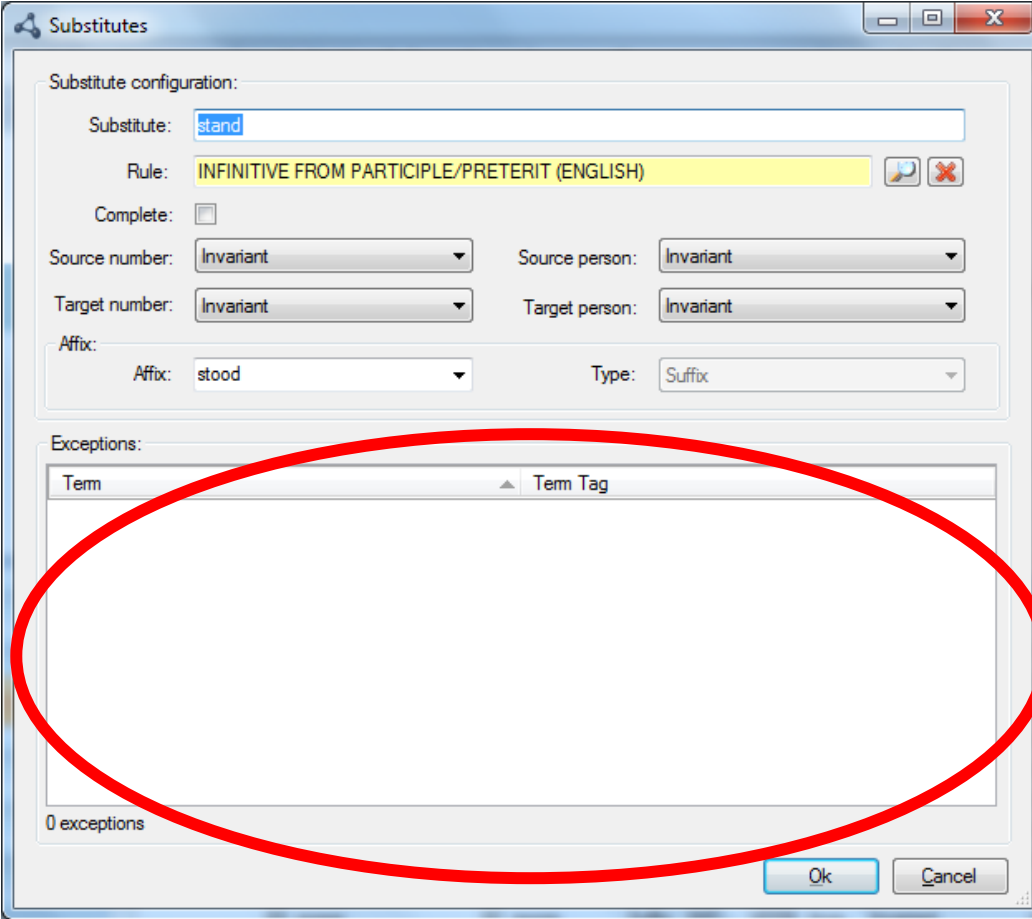
Normalization (Stemming)



Normalization to Plural

- The Normalization rule should be “nothing” to “s” or “es” or...
- No support to “empty” suffix
- So particular rules should be created
- Eg. For Interface -> Suffix “ace” to “aces”
- Etc.

Normalization (Stemming) - Exceptions



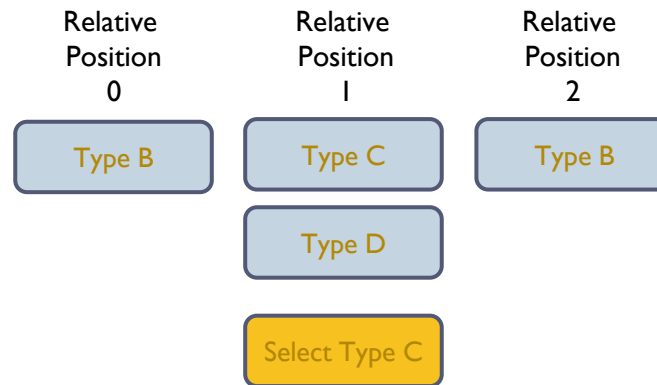
- Exceptions are applied to terms before the affix work
- If a term is the exception => do not apply the substitution.
- For Latin Languages, Exceptions are also applied after reflective word separations
 - Eg. Retirarse –
 - Retirar SE

Disambiguation / Tagging

- Process needed to solve multiple candidates created by the normalization process.
- Based on Context
 - At both sides of the ambiguous term
- Based on Disambiguation rules + Bigrams Statistics + Rules of thumb
- Application Order:
 - Disambiguation rules
 - Bigrams Statistics
 - Rules of Thumb
 - The tag of the candidate that matches with the input text (if not unclassified)
 - The tag with more probability of occurrence by itself
 - The Noun tag if a candidate is of that type
 - The tag of the first candidate

Disambiguation / Tagging : (Rules)

➤ Disambiguation rules

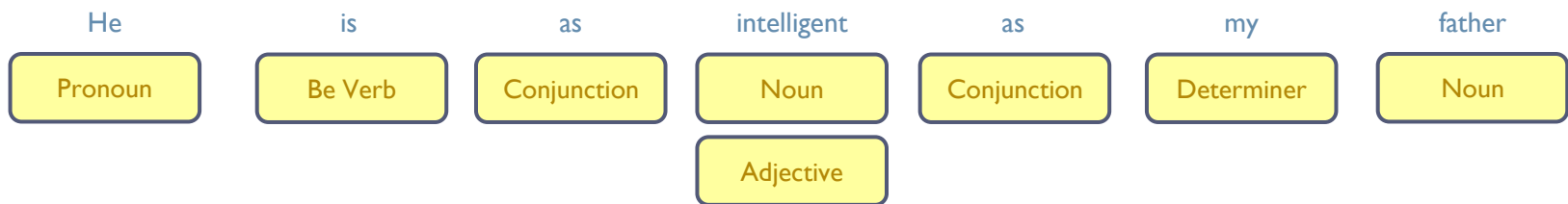


- If In Position 0 we have a candidate of type B and in Position 1 we have a candidate of Type C and a Candidate of Type D and in position 2 we have a candidate of type B => select the candidate of type C in position 1 as the right one

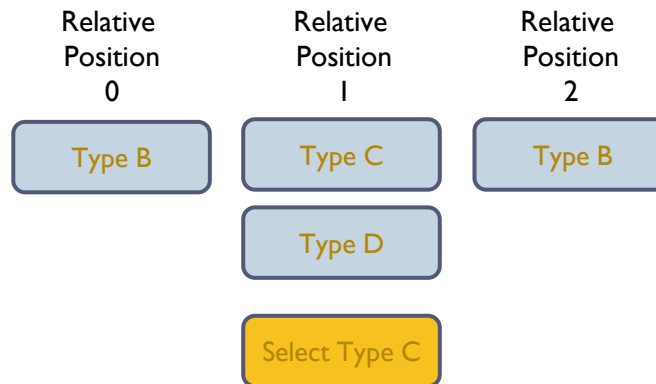
Disambiguation / Tagging : (Rules)

➤ Example:

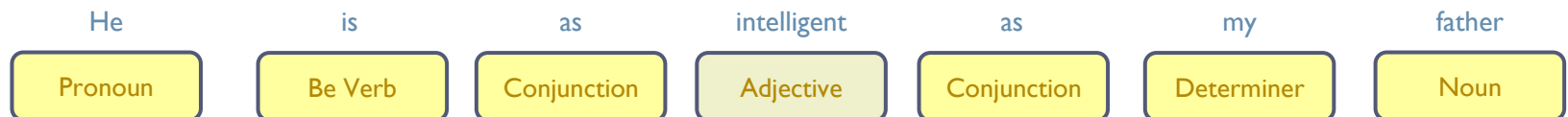
Initial Sentence



Rule

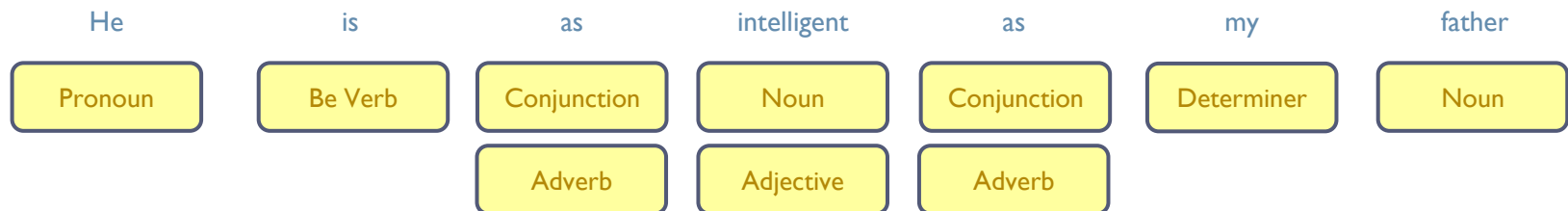


Final Sentence



Disambiguation / Tagging : (Rules)

- Rules are applied in Order of weight
- All the valid candidates at the moment of the application of the rule are considered
 - Which can imply possible problems
- Eg. If the sentence is:



- And the rule of the previous slide is applied first => the rest of the rules should work with the following sentence:



Disambiguation / Tagging : (Rules)

knowledgeMANAGER

By The REUSE Company

Domain Management Subdomains Indexing & Search Documents Configuration Save Exit

Tagging

Qualifications

Tagging rules

Bigram disambiguation

Tag probabilities

Terminology

Tokenization

Stemming

Tagging

Light Ontology

Inference rules

Clusterization

Domain Management > Tagging > Tagging rules

Search fields:

Description:

Enabled: ☐

Search

Tagging rules:

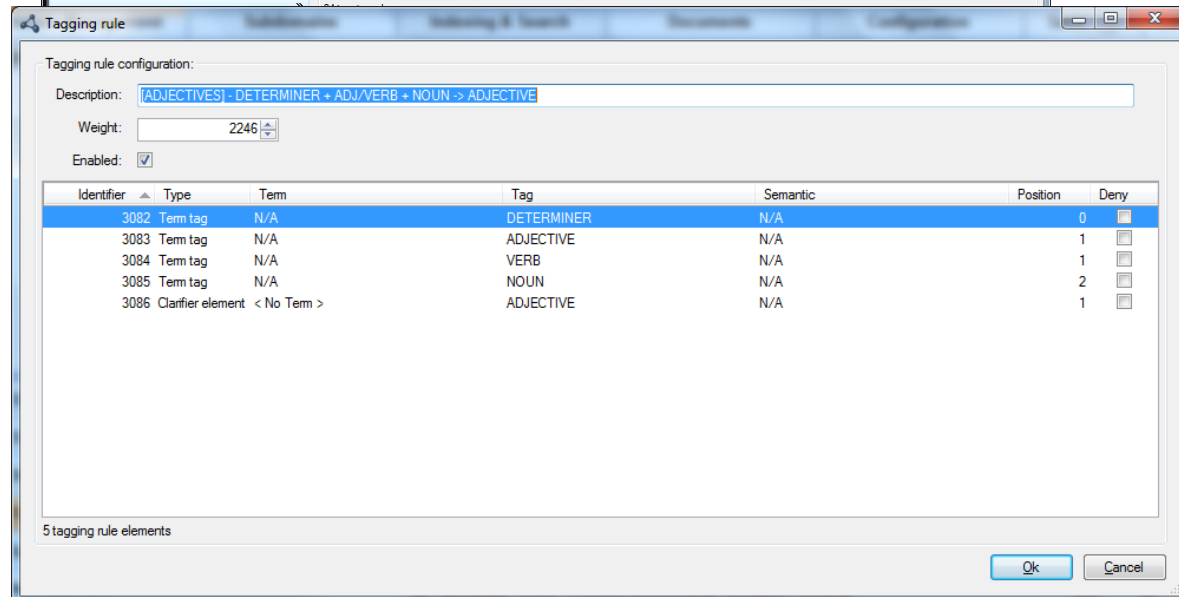
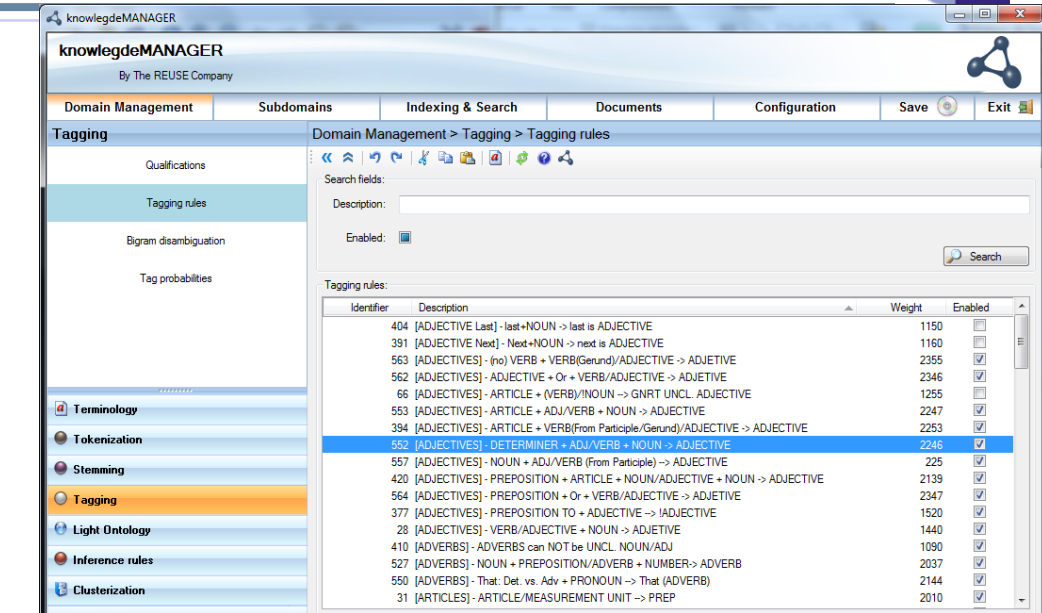
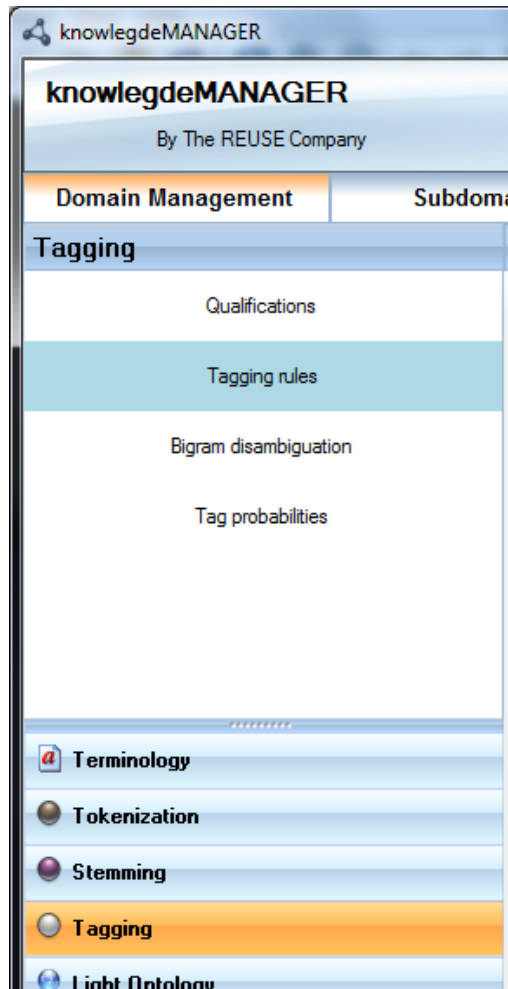
Identifier	Description	Weight	Enabled
404	[ADJECTIVE Last] - last+NOUN -> last is ADJECTIVE	1150	<input type="checkbox"/>
391	[ADJECTIVE Next] - Next+NOUN -> next is ADJECTIVE	1160	<input type="checkbox"/>
563	[ADJECTIVES] - (no) VERB + VERB(Gerund)/ADJECTIVE -> ADJECTIVE	2347	<input checked="" type="checkbox"/>
562	[ADJECTIVES] - ADJECTIVE + Or + VERB/ADJECTIVE -> ADJECTIVE	2346	<input checked="" type="checkbox"/>
66	[ADJECTIVES] - ARTICLE + (VERB)/NOUN -> GNRT UNCL. ADJECTIVE	1255	<input type="checkbox"/>
553	[ADJECTIVES] - ARTICLE + ADJ/VERB + NOUN -> ADJECTIVE	2247	<input checked="" type="checkbox"/>
394	[ADJECTIVES] - ARTICLE + VERB(From Participle/Gerund)/ADJECTIVE -> ADJECTIVE	2253	<input checked="" type="checkbox"/>
552	[ADJECTIVES] - DETERMINER + ADJ/VERB + NOUN -> ADJECTIVE	2246	<input checked="" type="checkbox"/>
557	[ADJECTIVES] - NOUN + ADJ/VERB (From Participle) -> ADJECTIVE	225	<input checked="" type="checkbox"/>
420	[ADJECTIVES] - PREPOSITION + ARTICLE + NOUN/ADJECTIVE + NOUN -> ADJECTIVE	2139	<input checked="" type="checkbox"/>
377	[ADJECTIVES] - PREPOSITION TO + ADJECTIVE -> ADJECTIVE	1520	<input checked="" type="checkbox"/>
28	[ADJECTIVES] - VERB/ADJECTIVE + NOUN -> ADJECTIVE	1440	<input checked="" type="checkbox"/>
410	[ADVERBS] - ADVERBS can NOT be UNCL. NOUN/ADJ	1090	<input checked="" type="checkbox"/>
527	[ADVERBS] - NOUN + PREPOSITION/ADVERB + NUMBER-> ADVERB	2037	<input checked="" type="checkbox"/>
550	[ADVERBS] - That: Det. vs. Adv + PRONOUN -> That (ADVERB)	2144	<input checked="" type="checkbox"/>
31	[ARTICLES] - ARTICLE/MEASUREMENT UNIT -> PREP	2010	<input checked="" type="checkbox"/>
399	[DETERMINER - CUANTIFIER] - CUANTIFIER DETERM.+ NOUN -> CUANTIFIER DETERM.	1200	<input type="checkbox"/>

83 tagging rules

Ready

Connected to 'D:\TRC - Repository\RQS-Docummentation\Products And Services\7-Events\2013 06 12 SNECMA\BDs\OOTB - V61.mdb'

Tagging : (Rules)



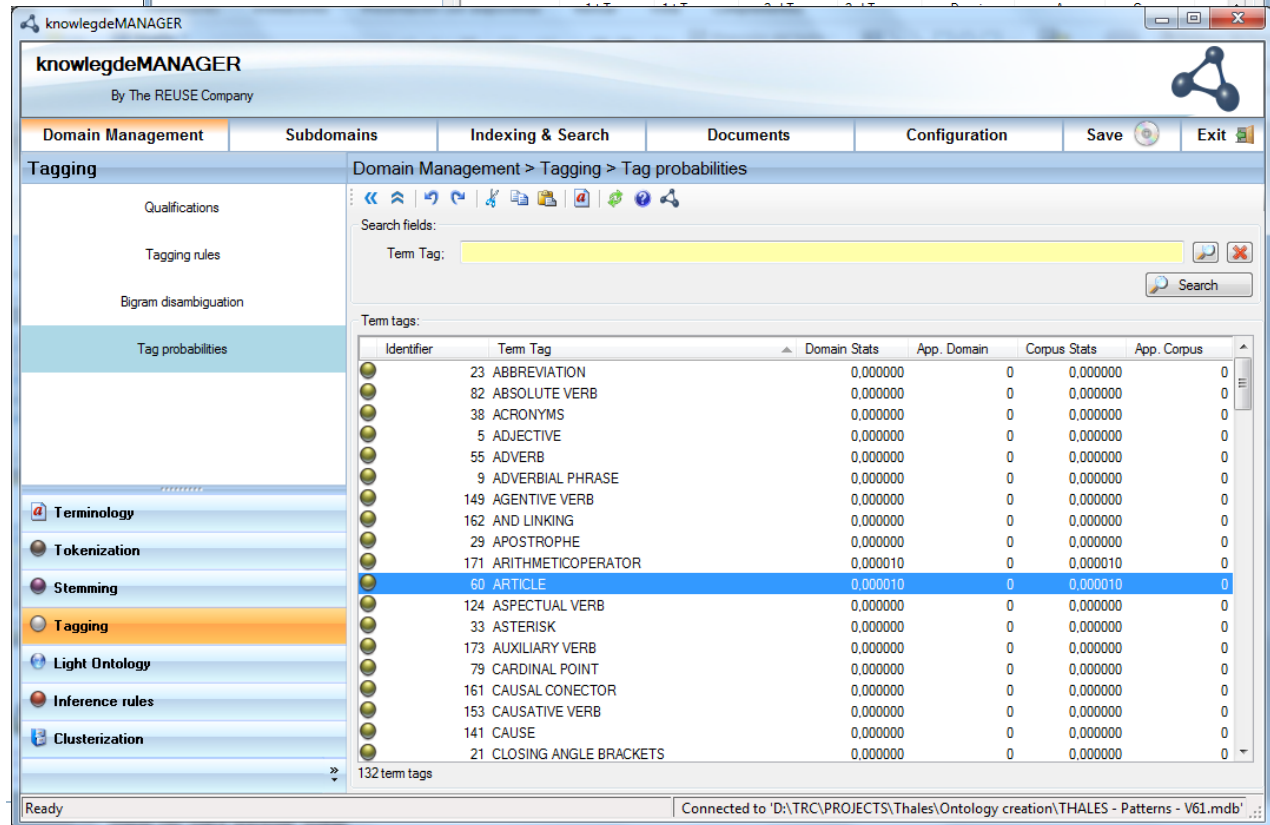
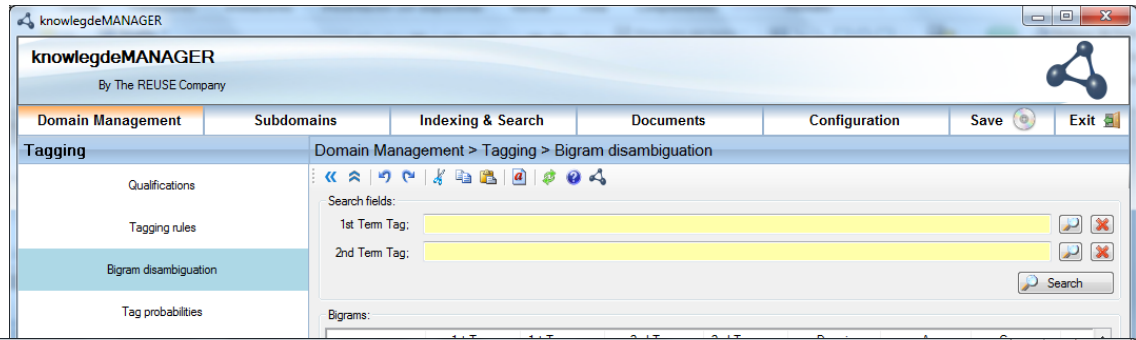
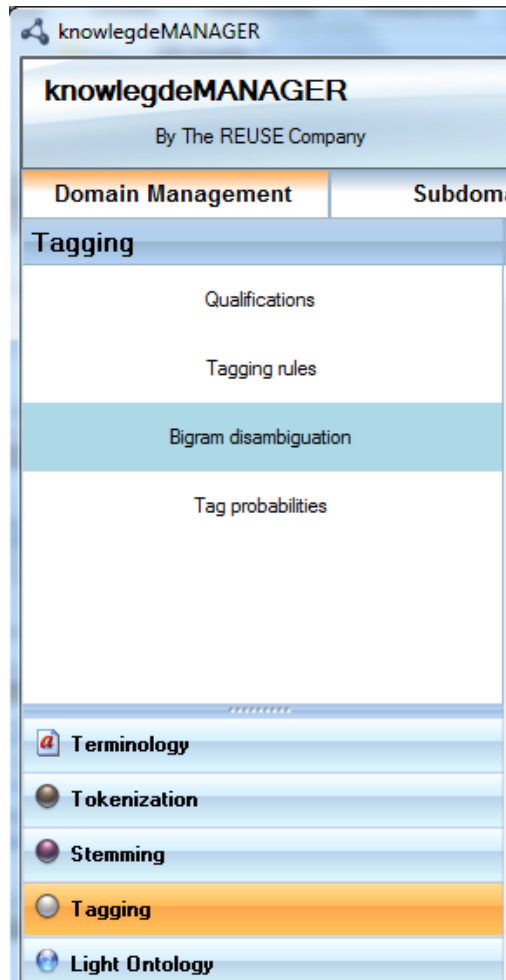
Disambiguation / Tagging : (Bigrams)

- Eg. If the sentence is:



- The following statistics will be checked:
 - Be Verb – Conjunction
 - Be Verb – Adverb
 - Conjunction – Adjective
 - Adverb – Adjective
- To decide which is the right tag for the first “AS”

Tagging : (Bigrams)





The
Reuse
COMPANY
CONFINIA



Innovation and Knowledge

The Reuse Company Trace > Reuse > Quality



Margarita Salas, 16 2nd Floor
Innovation Center
LEGATEC Technology Park
28919 Leganés – Madrid
SPAIN – EU



Tel: (+34) 912 17 25 96
Fax: (+34) 916 80 98 26



<http://www.reusecompany.com>



[@ReuseCompany](https://twitter.com/ReuseCompany)



contact@reusecompany.com