

A test of five terminology interfaces for SNOMED CT in clinical practice

Arabella D'Havé, Chief of Terminologie, Classification, Grouping & Audit
Dr. Ingrid Mertens, Head of Department Datamanagement

Belgian Federal Public Service of Health, Food Chain Safety and Environment
Terminology Center

ABSTRACT

Objective: The primary objective of this study is to evaluate the capability of different terminology interfaces to find the preferred term in the Belgian SNOMED CT subset for clinical findings and procedures through the use of free text labels in Dutch and French as used by the physicians in the Electronic Health record (EHR).

Design: All vendors or organisations offered a terminology server which can be integrated in electronic medical records as a terminology component for the retrieval of preferred terms based on the entry of free EHR terms. All systems were prototypes which were mounted for free. Testers from four Belgian hospitals tested the retrieval of a SNOMED CT concepts based on an input of at least 50 EHR labels and with at least two different user interfaces. Depending on its ranking in the returned results, the label was categorized.

Results: All providers have room to improve syntactic and semantic features.

All providers were keen to learn more from this study and to develop and improve the needed terminology interface functions. Only one interface (Indizen) appeared to be acceptable for use in practice with an average score of 6.7 on a scale of 10. The other interfaces were slightly below 5. The syntactic features available in all interfaces was 'exact match' search and 'word permutation'. All other syntactic and semantic functions were more or less present but could be largely improved.

Conclusion: The participation of terminology server providers was good despite the difficult context (test project without budget, integration of a new terminology in two languages, many testers all over the country). Several testers asked for a function to visualise the SNOMED CT hierarchy for navigation to drill down to the full description and cross maps with other coding systems.

METHODS

Six terminology interfaces were selected from the following vendors or organisations : Mondeca, Indizen, InterSystems, NICTIZ, Collibra, Linkbase. All participants offered a terminology server which can be integrated in electronic medical records as a terminology component for the retrieval of preferred terms based on the entry of free EHR terms. All systems were prototypes which were mounted for free. Finally Collibra withdrew from the study and didn't provide a test server.

Indizen and Mondeca were tested by all testers in all languages (French and Dutch). Intersystems, Linkbase and NICTIZ were tested only by the Dutch speaking testers.

In 4 Belgian hospitals, EHR labels were selected by testers and pasted into a terminology interface. A tester tested at least 50 EHR labels with at least two different test interfaces. Different testers tested different batches of EHR labels.

The results of the search were coded as follows:

- '0' Right Preferred Term is not found within the first 40 results
- '1' Right Preferred Term is found within the first 9 results (Preferred Terms)
- '2' Right Preferred Term is found between the 10th and 19th result (Preferred Terms)
- '3' Right Preferred Term is between the 20th and 39th result (Preferred Terms)

With regards to the search result, the tester filled out an evaluation sheet for the following syntactic and semantic features:

Syntactic / Semantic features	Evaluation
Fuzzy search	Is the interface capable to find the right Preferred term, even when some characters of the 'search term' are different from the Preferred term?
Exact match	If you enter the exact term (from the column 'Preferred term'), does the interface give you the Term on the first place of the results?
Word Decomposition	When the label is spelled in one word but the preferred term is spelled in more than one word: is the interface capable of finding the right word? Eg longkanker <-> long kanker
Word Permutation	When the words of the label are spelled in a different sequence than the preferred term: is the interface capable of finding the right word? Eg acuut longoedeem <-> longoedeem acuut
Spelling correction	When a word is slightly misspelled, is the interface capable of finding the right word? Eg infract <-> infarct
Phonetic equivalences	When a word is slightly misspelled with phonetic equivalence, is the interface capable of finding the right word? Eg pharyngitis <-> faryngitis
Semantic search (synonyms)	When the label is a common synonym is the interface capable of finding the right word? Eg cholelithiase <-> galstenen
Semantic search (parents & children)	Does the interface show the parents and children in a hierarchical tree? Or in a flat result list expanding automatically parent and children?
Predictive result	Can you understand the logic of the order of the results to adapt your search terms?

Result was converted to a score

- 'No' (Function is not available); 'Poor'; 'Moderate'; 'Good'

In the analysis the median score of all testers was used.

RESULTS

Number of labels tested per interface:

Interface	Fr	NI	Grand Total
Indizen	401	393	794
Intersystems		371	371
Linkbase		124	124
Mondeca	400	323	723
NICTIZ		373	373
Grand Total	801	1.584	2.385

There were 265 EHR labels scored by two different testers with the same interface. The percentage of preferred terms found per user interface:

	Total	1	2	3	Corrected Including exact match PT
Indizen	62%	58%	3%	2%	77%
Intersystems	39%	32%	5%	1%	54%
Linkbase	59%	55%	3%	1%	74%
Mondeca	43%	42%	1%	0%	58%
NICTIZ	11%	11%	0%	0%	26%
Grand Total	44%	41%	2%	1%	59%

Results on global evaluation:

Interface	Number evaluations	Average	Min	Max
Indizen	15	6,7	1	8
Intersystems	7	4,1	2	7
Linkbase	1	5,0		
Mondeca	15	4,9	3	8,5
NICTIZ	7	4,5	2	8

Results on syntactic criteria:

	Fuzzy map	Exact match	Word Permutation	Spelling correction	Phonetic equivalence
Indizen	good	good	good	poor	moderate
Linkbase	moderate	good	good	good	good
Mondeca	poor	good	good	good	good
Intersystems	poor	moderate	good	poor	poor
NICTIZ	no	good	good	no	poor

Results on semantic criteria:

	Word Decomposition	Semantic search (synonyms)	Semantic search (parents & children)	Predictive result
Indizen	moderate	good	no	poor
Linkbase	poor	good	no	poor
Mondeca	poor	moderate	no	poor
Intersystems	good	moderate	no	poor
NICTIZ	poor	no	no	poor

In this study the predictability of the results was considered as 'poor' which means that the physician has to go through the list of results causing loss of time. This is somewhat in contradiction with the fact that > 90 % of the found Preferred Terms appeared in the top ten of the results.

All interfaces gave a good result when the exact term was used.

Fuzzy map was considered as 'good' for Indizen and as 'poor' for the other interfaces.

All interfaces were capable to permute words within the Preferred term.

Spelling correction and phonetic equivalence was poor in all commercial interfaces except for Mondeca.

Word Decomposition was the most important missing feature for Preferred terms 'not found' in the test with Indizen. Mondeca had a good score.

Linkbase had a good score for semantic search. However Indizen had an equally good score using the submitted synonyms and SNOMED CT ontology.

CONCLUSIONS

In this study the predictability of the results was considered as 'poor' which means that the physician has to go through the list of results causing loss of time. This is somewhat in contradiction with the fact that > 90 % of the found Preferred Terms appeared in the top ten of the results.

Several testers asked for a function to visualise the SNOMED CT hierarchy for navigation to drill down to the full description and cross maps with other coding systems.

The needs of the physician or coder vary according to the context. A terminology interface should be capable to adapt his behavior accordingly.

A full semantic ontology representing a semantic network linked to SNOMED CT concepts is expected to convert natural language to the right Preferred Terms.

It is unclear to what extent hierarchical browsing will be used by the average physician.

Terminology interfaces should organise post-coordination in an easy way. The refinement with sub-concepts (qualifiers) for the given primary concepts should pop up automatically and should allow a fast selection.

AKNOWLEDGEMENTS

CHU Citadelle Liège

CHU ULB Bruxelles

CHU Charleroi

Universitaire Ziekenhuizen Antwerpen (UZA)

