MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Evaluating Bilingual Medical Terminologies with Word Alignment Methods

Mikael Nyström[a], Magnus Merkel[b], Håkan Petersson[a], Hans Åhlfeldt[a]

[a]*Department of Biomedical Engineering, Linköping University, Sweden*
[b]*Department of Computer and Information Science, Linköping University, Sweden*

Abstract

Automatic word alignment in parallel texts can be used to generate lexicons, but the lexicons' quality is depending on the quality and type of the input resources. In this study the English and Swedish versions of the medical terminology systems ICD-10, ICF, NCSP, KSH97-P and MeSH are automatically word aligned and the influence the different systems have on the quality is evaluated. The most generally useful resources were generated from ICD-10 and the size of the partitions used to generate the resources only partly affects the quality.

Keywords:

classification, linguistics, medical informatics computing, terminology, medical terminology systems, word alignment

Introduction

One solution to cut costs for manual translations of medical terminology systems are to reuse already translated systems. In word alignment the task is to find correspondences on the word or phrase level between a source text and its translation, and the result can be reused in new translations. In an earlier study we word aligned a union of ICD-10, ICF, MeSH, NCSP and KSH97-P [1]. In this study we explore the influence the different terminology systems have on the recall and precision by varying how the systems are utilized in word alignment.

Materials and methods

Alignment tools

The ITools suite is used in this study for the word alignment [2]. First the user creates word alignment training data in an interactive environment. Then automatic alignment is done in a primarily linguistically oriented manner where different types of resources are combined in a voting procedure to select the best possible alignments. The static resources used are dictionaries and pattern resources for common and unwanted parts-of-speech correspondences and standard punctuation. The dynamic resources used are generated during the training stage. The statistic resources used are dictionaries produced using statistical analysis on the parallel texts.

Used terminologies and terminology partitions

The rubrics of the medical terminology systems ICD-10, ICF, MeSH, NCSP and KSH97-P that exist in parallel in both English and Swedish have been extracted for use in the study. A manual exploration divided the terminology systems into partitions with similar content and ordered the partitions from partition 1 with the highest assumed quality to partition 8 with the lowest. The exploration found the partitions 1: MeSH one word in English or Swedish rubric (13,514 rubrics), 2: MeSH more than one word in English and Swedish rubric (5,568 rubrics), 3: ICF (1,496 rubrics), 4: KSH97-P (968 rubrics), 5: ICD-10 except chapter 2 level 4 (10,791 rubrics), 6: NCSP except chapter N (4,137 rubrics), 7: ICD-10 chapter 2 level 4 (713 rubrics), 8: NCSP chapter N (1,388 rubrics). Partition 1 was used as a static dictionary for the alignment of partition 2–8.

Preparation and manual word alignment

Statistical resources were generated from each of the partitions 2–8. Then from each of the partitions 2–8, 5 % were randomly sampled to a training set and 5 % to a test set. These sets were interactively aligned by one of the authors following a style guide. The training sets were used to build training resources and the test sets as gold standards for evaluating the automatic word alignment's recall, precision and F-score.

Automatic word alignment

The test sets were automatically word aligned and the partitions used for building the statistic and dynamic resources and to automatically word align were altered according to Table 1. The main assumption to examine in batch 1 was similarities and differences inside each partition, in batch 2 the similarities and differences between a partition and the union of all partitions, in batch 3 if the performance of the automatic word alignment improves monotonously when resources from more and more partitions are used or if some partitions worsen the performance and in batch 4 the similarities and differences between the union of all partitions and a partition.

Results

The results of the word alignment are presented below.

Table 1 - Word alignment configurations and results

Batch	Run	Resource partition	Align partition	Recall	Precision	F-score
1	1	2	2	0.72	0.70	0.71
	2	3	3	0.80	0.80	0.80
	3	4	4	0.71	0.68	0.69
	4	5	5	0.80	0.78	0.79
	5	6	6	0.83	0.78	0.80
	6	7	7	0.63	0.58	0.60
	7	8	8	0.84	0.85	0.84
2	1	2	2–8	0.65	0.65	0.65
	2	3	2–8	0.62	0.63	0.62
	3	4	2–8	0.64	0.64	0.64
	4	5	2–8	0.75	0.73	0.74
	5	6	2–8	0.67	0.67	0.67
	6	7	2–8	0.64	0.65	0.64
	7	8	2–8	0.64	0.65	0.64
3	1	2	2–8	0.65	0.65	0.65
	2	2–3	2–8	0.67	0.66	0.66
	3	2–4	2–8	0.69	0.67	0.68
	4	2–5	2–8	0.77	0.74	0.75
	5	2–6	2–8	0.80	0.76	0.78
	6	2–7	2–8	0.80	0.77	0.78
	7	2–8	2–8	0.81	0.77	0.79
4	1	2–8	2	0.73	0.71	0.72
	2	2–8	3	0.81	0.81	0.81
	3	2–8	4	0.81	0.76	0.78
	4	2–8	5	0.81	0.78	0.79
	5	2–8	6	0.84	0.79	0.81
	6	2–8	7	0.74	0.65	0.69
	7	2–8	8	0.85	0.85	0.85

Discussion

In batch 2, all partitions except partition 5 have quite a similar recall and precision despite their varying sizes. This indicates that the sizes of the partitions used to generate the resources only partly influence the alignment results.

In batch 3, the recall and precision increase or remain unchanged for each added partition. This implies that none of the added resources decreases the word alignment quality.

Batch 4 has better results than the other batches, but compared to batch 1 the improvements of using resources generated from all partitions are in general small.

Partition 2 gives not as good results as most of the other partitions in batch 1 and 4, which indicates that there might be content-related aspects that make the alignment difficult.

The best resources are generated from partition 5, and this partition seems to cover the medical domain, represented as the union of all partitions, best of the single partitions. This is probably because it contains the highest number of rubrics of the partitions. This result is interesting because the partition contains ICD-10 and most of the other included terminology systems cover other sub-domains of medicine than diseases. Resources from partition 5 show in general better result than partition 4. Both partitions have similar content and structure, but partition 5 is eleven times larger than partition 4. The large difference in size and the moderate difference in results show that only adding larger resource partitions can be an expensive way to improve the results.

Partition 7 is difficult to align according to the results of batch 1 and 4. This can probably be explained by structural and semantic differences between the Swedish and English versions.

Conclusions

A larger resource partition generates resources that only to some extent have higher quality, so only adding larger resource partitions can be an expensive way to improve the results. ICD-10 seems to cover the medical domain best of the included terminology systems and generated the best general resources. Systematic differences in structure between rubrics and their translations make the rubrics harder to align.

Acknowledgments

This work was performed in the framework of the EU-funded Network of Excellence entitled Semantic Interoperability and Data Mining in Medicine (SemanticMining).

References

[1] Nyström M, Merkel M, Ahrenberg L, Zweigenbaum P, Petersson H, Åhlfeldt H. Creating a medical English-Swedish dictionary using interactive word alignment. BMC Med Inform Decis Mak. 2006 October 12;6(35).

[2] Foo J, Merkel M. Building standardized term bases through automated term extraction and advanced editing tools. To be published in the Proceedings of the International Conference on Terminology; 2006 November 16-17; Antwerp; 2006.

Address for correspondence

Mikael Nyström, Department of Biomedical Engineering, Linköping University, SE-581 85 Linköping, Sweden. mikny@imt.liu.se.