



How can we use clinical corpora to assist the clinician, her managers and clinical research?

Dr. Hercules Dalianis

Department of Computer and Systems Sciences (DSV)
Stockholm University, Sweden

Sala de Grados • 9 de mayo de 2011 • 17:00
entrada libre hasta completar el aforo

resumen:

Today a large number of Electronic Patient Records (EPRs) are produced for legal reasons but they are never reused neither for clinical research nor for business (hospital) intelligence reasons. Moreover, it is also alarming that the clinician's daily work in documenting the patient status is rarely supported in a proper way. We are aiming to change these facts. Clinical corpora form an abundant source to extract valuable information that can be used for this purpose. The Stockholm EPR Corpus is a huge clinical corpus written in Swedish, containing over one million patient records distributed over 800 clinics encompassing three years from the Stockholm area. We have explored subsets of this corpus with the aim of understanding the whole corpus and its domain(s). In one experiment we annotated a subset of the corpus for de-identification, and we created a gold standard for training and evaluation of automatic de-identification tools. In another experiment we investigated the relations of diagnosis codes (ICD-10) for co-morbidity analyses and found interesting results. We have also developed a method for automatic support in assigning new ICD-10 codes on newly entered clinical text, but also for evaluating already assigned ICD-10 codes. Finally we have tried to understand what exactly is written in the corpora, with the aim to construct information extraction tools that can distinguish between the factuality of diagnoses. Is the diagnosis certain, negated, or uncertain to some extent? Two annotators with clinical background have annotated a subset of the corpus for factuality levels.

sobre Hercules Dalianis:

Hercules Dalianis received his Ph.D in 1996. He was a post doc researcher at University of Southern California/ISI in Los Angeles 1997-98. He held a three-year guest professorship at CST, University of Copenhagen during 2002-2005, founded by Norfa, the Nordic council. Dalianis works in the interface between university and industry with the aim to make research results useful for society. He has specialized in the area of human language technology, to make computer to understand and process human language text, but also to make a computer to produce text automatically. Examples on applications are automatic text summarization and search engines with built in human language technology support as for example stemming, spell checking, compound splitting to improve the information extraction. Currently he works in the area of text mining and medical informatics focused on electronic health records. He has more than 20 years of experience of his research area and he has been project leader and received funding for over 15 national, Nordic and European research projects