INLP-MAI Term 2014-2015, Fall semester Final exam

The medical domain is one of the most challenging for textual NLP but also it is one of the domains that has attracted more attention to the NLP research and development communities.

Roughly two kind of texts are usually used: medical literature (papers, books, ...) and medical reports. The later are clearly more challenging. Million of medical reports are daily produced by different medical professionals (doctors, nurses, radiologists, ...) in different steps of health treatment.

Different types of medical reports differ in their length, degree of formality, terminology used, ... We will focus on radiology reports that use to be short texts resulting from radiology explorations (X rays, scanning, tomography, ...) that describe the image(s) resulting from the exploration. Consider the following examples:

1) *On top of left arm side area in the subcutaneous cell, we observe small rounded shaped huts with heterogeneous calcifications. Measures 1.9 x 1.2 x 2.7 cm in diameter. Find seems compatible with BCGitis*.

2) *Liver: normal size and ecoestructure. Hepatic artery, portal vein and suprahepatics without alterations. Intra- and extrahepatic biliary not extensive. Gallbladder: acalculous. Pancreas: ecoestructure and size normal. Spleen: normal size and ecoestructure. Longitudinal diameter: 8 (cm) retroperitoneum vascular without alterations. No adenomegalias detected.*

3) *Static study: both hips centered. Osteocartilaginosis adequate coverage. Left femoral nuclei with incipient ossification. Dynamic study both hips stable. CONCLUSION: ultrasound characteristics of normal hips.*

4) *A magnetic resonance imaging study will be scheduled as an outpatient later to rule out a small vascular malformation.*

Answer the following five questions. All the questions are equally weighted (2 points each):

1. Characterize the sublanguage used in radiology reports. Comment briefly the problems (compared with general texts) that occurs in this genre of documents.

2. List the kind of information and knowledge that could be extracted from these documents. Use example 1) above for illustrating your answer.

3. How to deal with the information extraction tasks presented in 2)? Which linguistic tasks are involved? Describe the tasks focusing on the challenges derived from the characteristics presented in 1).

4. Which knowledge sources are needed for 3)? How could be obtained? Quantify roughly the size of these knowledge sources.

5. Discuss about a multilingual setting. Could some of the resources in 4) for a language be used (perhaps with some limited human intervention) for other languages.