Slide 1:

Hello…

In my project I am trying to extract certain information out of clinical practice guidelines with the aim of translating them into the Asbru language.

Slide 2:

The reasons for this are as obvious as simple.

These guidelines contain information about clinical processes. But since the guidelines are consisting of unstructured plain-text, they cannot be automatically processed by machines, they are not executable.

Therefore they have to be translated into a formal representation like the Asgaard project, respectively the Asbru language. The Asgaard supports the design and the execution of skeletal plans by a human executing agent.

It is a task-specific and intention-based plan representation language, designed for management-task plans.

To be more precisely I am just dealing with information contained in conditions.

Before translating conditions of a CPG into Asbru, they have to me BPMN modeled into a semi-structured format, to make it easier to find the phrases that contain the conditions. Automated finding of conditions in complete plain-text documents would be a very complex task by its own.

Slide 3:

Since Information Extraction is an important task of information technologies since many years, there are many different approaches for information extraction in general, like using natural language processing, ontologies, support vector machine, heuristic methods. Discussing them would take too much time.

The most important work, when speaking of IE of CPG with the aim of importing the information into the Asbru language is done at the University of Technology Vienna itself, where the unified medical language system (UMLS) is used as medical terminology system and the many-headed-bridge as a format. The MHB format uses xml for abstracting CPG while translating it into other languages. It provides chunks of information which correspond to chunks of the natural language text and splits them into 8 different groups that cover the medical concepts. Then the chosen medical concepts are linked to a structure with the aim of recreating the logical relations represented by the guidelines, using the UMLS semantic network. The schema of the structured document is determined by the target system. To identify the information a rule based system is used on both syntactic and semantic information.

Another approach is given by a university in Mexico.

IE systems can be classified based on two approaches.

**Knowledge Engineering (KE):** This is focused on an empiric method or based on a

domain corpus to develop efficient and robust Natural Language Processing systems

**Machine Learning (ML):** This has a well-known set of documents and outputs and uses

a set of patterns to extract knowledge by means of Machine Learning techniques

They used the machine learning approach and built a three level algorithm based on 3 heuristic patterns and created a rule based extraction system.

The first is the phrase pattern level, also called lexical level, where a lexical parser is used to identify the relevant information.

The sentence pattern or syntactic level parses the whole document and splits it into sentences. The sentences are processed with regard to its context which is obtained by captions.

In the speech pattern level semantic aspects are solved and the design and structure of the final document are improved. In this level the sentences and actions are categorized and their relations are set.

Slide 4:

First step is to read the input file, that contains the BPMN modeled CPG as xml. With the GATE embedded I used ANNIE for IE and with the MetaMap-plug-in I communicated with the 2 servers to process the information with the medical terminology system UMLS.

I implemented a codition analyzer that tries to evaluate the results given by the MetaMap and classifies them regarding to the corresponding Asbru condition.

For processing the condition structure I implemented the visitor pattern, which is among others used by my translator – that translates all found conditions into the Asbru language.  
The result is a document in xmlj containing all conditions in the Asbru schema.

Slide 5:

I used Java and Eclipse as development environment. Because this is my first experience with Java, I skipped concepts like unit testing, I just implemented a small console application and debugging for testing purposes.

As already mentioned input and output of the plug-in is each just one file.

For Information extraction I used GATE embedded, which is the java api of the GATE. It also includes a plug-in that allows the usage of the MetaMap Java API, that communicates with 2 servers (I installed them both locally which is not necessary) that use the UMLS to identify medical concepts in texts. With an implementation of the visitor pattern I processed the arised condition structure and created the Asbru conditions.

Slide 6:

This is an example of a conditionExpression tag contained in a BPMN modeled guideline. As you can see the conditions are still in plain-text. As you can see there are three conditions linked with a logical or.

Slide 7:

After I have processed it, the conditions are translated into the Asbru language and look like this.

They correspond to the Asbru “simple-condition”, where they are combined with an “Or”. I just formulated the half of the first condition, which expresses that the age is higher than 65.

As you can see, to reach a format that can be auto-processed it needs a lot more structure compared to the plain-text version.

Slide 8:

Although I am not finished yet, the biggest weakness will definitely be the small range of conditions that can be processed.

It is very difficult and tricky to correct interpret the results given by the MetaMap.

My work for the next weeks will be to improve the algorithm and to make it possible to detect and translate some more conditions.

An aim for the future would be to translate all conditions, but I doubt that all the semantics will be recognized in the near future.

I am unfamiliar with Java and Eclipse, I prefer the .net world.

Gate embedded is easy to include, to set up and to use.

If you want to use MetaMap on your local machine, you need to use 2 servers. There is a MetaMap Java-API use the servers and a gate-plug-in for MetaMap that uses it.

Slide 9:

Thanks. If there’re any questions feel free to ask.