A1.1 BRITISH NATIONAL CORPUS

BNC is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century (University of Oxford 2013).

The written part of the BNC (90%) includes, for example, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. It includes many different styles and varieties, and is not limited to any particular subject field, genre or register.

For written sources, samples of 45,000 words are taken from various parts of single-author texts. Shorter texts up to a maximum of 45,000 words, or multi author texts such as magazines and newspapers are included in full. Sampling allows for a wider coverage of texts within the 100 million limit and avoids over representing idiosyncratic texts.

A1.2 WORDNET

Wordnet is a lexical database for the English language. It groups English words into sets of synonyms called synsets, provides short, general
definitions and records the various semantic relations between these synonym sets. The purpose is to produce a combination of dictionary and thesaurus that is more intuitively usable and to support automatic text analysis and artificial intelligence applications (Princeton University 2013). The database and software tools can be downloaded and used freely. The database can also be browsed online. Wordnet is created and maintained by the Cognitive Science Laboratory of Princeton University. The database contains 155,287 words organized in 117,659 synsets for a total of 206,941 word-sense pairs.

It distinguishes between nouns, verbs, adjectives and adverbs as they follow different grammatical rules. It does not include prepositions, determiners etc. Every synset contains a group of synonymous words or collocations and different senses of a word are in different synsets.

**A1.3 OPEN NLP**

The Apache OpenNLP library is a machine learning based toolkit for the processing of natural language text (The Apache Software Foundation 2013). It supports the most common NLP tasks. These tasks are usually required to build more advanced text processing services. OpenNLP also includes maximum entropy and perceptron based machine learning. It also provides a large number of pre-built models for a variety of languages, as well as the annotated text resources that those models are derived from.

The Apache OpenNLP library contains several components, enabling one to build a full NLP pipeline. These components include: sentence detector, tokenizer, name finder, document categorizer, POS tagger, chunker, parser, coreference resolution. Components contain parts which enable one to execute the respective NL task, to train a model and often also to evaluate a model. Each of these facilities is accessible via its application interface.
program interface. In addition, a command line interface is provided for convenience of experiments and training.

**A1.4 ARKREF**

ARKref is a tool for noun phrase coreference. ARKref belongs to a family of rule-based coreference systems that use rich syntactic and semantic information to make antecedent selection decisions (Carnegie Mellon University- Language Technologies Institute 2013). It is a deterministic, rule-based system which uses syntactic information from a constituent parser and semantic information from an entity recognition component, to constrain the set of possible NPs that could be antecedents for a given mention. It encodes both syntactic constraints as well as semantic. After filtering candidates with these constraints, it selects the antecedent, the candidate noun phrase with the shortest (cross-sentence) tree distance from the target. Antecedent decisions are aggregated with a transitive closure to create the final entity graph.

**A1.5 THE STANFORD PARSER**

A natural language parser is a program which works out the grammatical structure of sentences, for instance, which group of words go together and which words are the subject or the object of a verb. Stanford parser is a java implementation of probabilistic natural language parsers. It is a good performance statistical parsing system. A GUI (Graphical User Interface) is provided for viewing the phrase structure tree output of the parser.

**A1.6 DRUPAL**

Drupal is an open source content management platform, powering millions of websites and applications. It is built, used and supported by an
active and diverse community of more than 630,000 users and developers. It is distributed under the terms of the General Public License (Drupal 2013). The repository provides a large number of services, systematically classified in 47 categories. The Drupal project's principles encourage modularity, standards, collaboration, ease-of-use and more. Drupal is a framework for building dynamic web sites offering a broad range of features and services including user administration, publishing workflow, discussion capabilities, news aggregation and metadata functionalities, using controlled vocabularies and XML publishing for content sharing purposes.

A1.7 GR TOOL

GR Tool supports the forward and backward reasoning in Tropos. It is a graphical tool in which it is possible to draw the goal models and run the algorithms and tools for reasoning. The algorithms for the forward reasoning have been fully developed in java and are embedded in the GR Tool (Tropos 2013).

A1.8 BROWN CORPORA

The Brown Corpus was the first computer readable general corpus of texts prepared for linguistic research on modern English. It contains of over one million words of running text of edited English prose printed in the United States (Corpus resource database 2013).

A1.9 VERBNET

Verbnet is the largest online verb lexicon currently available for English. It is a hierarchical domain independent, broad coverage verb lexicon with mappings to other lexical resources such as wordnet, Xtag, and framenet (University of Colorado 2013). Verbnet is organized into verb
classes through refinement and addition of subclasses to achieve syntactic and semantic coherence among members of a class. Each verb class in verbnet is completely described by thematic roles, selectional restrictions on the arguments and frames consisting of a syntactic description and semantic predicates with a temporal function. Each class contains a set of syntactic descriptions, or syntactic frames, depicting the possible surface realizations of the argument structure for constructions such as transitive, intransitive and prepositional phrases.
This section provides a description of LCS and MSS.

A2.1 LIGHT CONTROL SYSTEM

LCS case study gives a description of the problem, “Light Control System” (Borger & Gotzhein 2000). The light control system is intended to develop a software that automatically controls the ON and OFF of lights in a building.

The requirements document includes the architecture of the building, functional needs, user needs, non functional needs and interface requirements. Some of the sample requirements are listed below.

- If a person occupies a room, there has to be safe illumination, if nothing else is desired by the chosen light scene.

- As long as the room is occupied, the chosen light scene has to be maintained.

- If the room is reoccupied within T1 minutes after the last person has left the room, the chosen light scene has to be reestablished.
• If the room is reoccupied after more than $T_1$ minutes since the last person has left the room, the default light scene has to be established.

• For each room, the chosen light scene can be set by using the room control panel.

• Before a person enters one hallway section from another one or from a staircase, the hallway section ceiling light group in the section being entered has to be on.

• Use daylight to achieve the desired light setting of each room and each hallway section whenever possible.

• If a malfunction occurs, the facility manager has to be informed.

• If a malfunction occurs, the control system supports the facility manager in finding the reason.

• The system provides reports on current and past energy consumption.

• If any outdoor light sensor does not work correctly, the default light scene for all rooms is that all ceiling light groups are on.

• If any outdoor light sensor does not work correctly and a hallway section is occupied, the ceiling light group in this hallway section has to be on.

• If any motion detector of a room or a hallway section does not work correctly, the control system should behave as if the room or the hallway section were occupied.
A2.2 MEETING SCHEDULER SYSTEM

This case study describes the requirements document used for MSS. The requirements document is developed at the University of Texas at Dallas (University of Texas 2013).

The MSS is a meeting scheduler intended for the usage in an enterprise. The MSS is a web based system that requires authentication for scheduling meeting. Scheduling a meeting involves diverse commitments and people from different background and with different preferences. A tool for scheduling a meeting provides a mechanism for better time planning and utilization.

The requirements document is intended to provide an overview of MSS and general aspects of the project. It articulates the specification of MSS. It encompasses related functional and non functional requirements, architecture, stakeholders, etc necessary for specifying the system. The document provides explanations of the system to be developed. It is set out in order with IEEE 830 standard with some modifications. It encompasses the resolution of each requirement inconsistency, ambiguity and other concerned issues. It also clarifies how each requirement is deduced and how requirements are inter-related. It encloses the process of analysis and refinement of requirements. Some of the sample requirements from the case study are listed below.

- The meeting initiator will ask all potential meeting attendees for set of dates they cannot attend the meeting (exclusion sets) and the set of dates they can attend the meeting (preference sets)
- A meeting date shall be defined by a (date, time) pair.
- The exclusion and preference set should be contained in some time interval described by the initiator.

- The meeting initiator could ask the active participants to provide any special equipment requirements on the meeting location.

- The initiator may also ask important participants to state preferences about the meeting location.

- It is absolutely necessary, however, to allow each meeting to take place in a virtual place.

- The system shall support conflict resolution according to the policies stated by the client.

- The amount of interaction among participants (e.g., number and length of messages, amount of negotiation required) should be kept minimal.

- The system should accommodate as much decentralized requests as possible and any authorized user should be able to request a meeting independently of her whereabouts.

- Physical constraints should not be broken, e.g., a person may not be at two different places at the same time and a meeting room may not be allocated to more than one meeting at the same time.