Ambiruptor: The Lexical Ambiguation Interruptor

Project Proposal

October 23, 2015

Presentation

The main objective is to produce an efficient tool that gives the correct meaning of an ambiguous text. Our tool will be based on several machine learning concepts. We are going to use annotated texts from the internet to train our tool and resolve ambiguities in texts.

Our team consists of 8 members: Boumediene Brikci Sid, Ievgeniia Oshurko, Maria Boritchev, Pierre Ohlmann, Samir Tendjaoui, Simon Mauras, Thi Xuan Vu and Victor Hublitz. The coordinators of the project are Simon Mauras and Ievgeniia Oshurko.

Contents

Motivation

Word Sense Disambiguation is a Natural Language Processing task that lies in assignment of appropriate meaning of the word according to the given context, and its separation from other possible meanings. Possible applications of Word Sense Disambiguation:

- Machine Translation.
- Information Retrieval.
- Semantic Parsing.
- Speech Synthesis and Recognition.

1 Objectives

The aim of our project is to develop a tool for disambiguation of text written in natural language. In order to do that we are going to:

- Mine data from existing sources.
- Preprocess data into training corpus.
- Create a tool using learning algorithms.
- Evaluate our model.
- Create user-friendly interfaces (if we have time).

2 Plan of Our Work

2.1 Research on the State of the Art

The first part of our work is going to be bibliographical search. On this stage we are going to split in 3 groups working on the following topics:

- Usual technics for text disambiguation.
- Machine learning algorithms.
- Existing tools for data mining.

The goal of the first topic is to find what has already been done and what is currently studied. The purpose of the second group is to compile a list of several machine learning technics that can be used to solve our problem. The third group is going to look for tools that would enable us to mine workable data from the internet. At the end of this part, each group is going to do a quick summary on their results, so that everyone is aware of each part.

2.2 Design

During the second part we are going to think about the detailed structure and the design of the software we want to implement.

2.3 Intermediate report

The intermediate report will clarify all the details of the upcoming implementation. A small group of two or three people will summarize the decisions made during the design part.

2.4 Implementation

The implementation will be divided between 2 groups:

- Adaptation of the chosen tools to extract workable data from the internet.
- Coding of the different modules (defined during the design part).

The progress of the second group depends on the results of the first group.

2.5 Testing

We will launch our tool on all the data we are able to process and it might take a while. Validity measures will be estimated.

2.6 Addons

As soon as the Ambiruptor tool becomes functional, the creation of user-friendly interfaces will be assigned to a small group of students.

- Web application.
- Pdf reader plugin.

We can expand this part if the project goes well and fast, as we can shorten it if we get stuck in one of the previous parts.

2.7 Final report

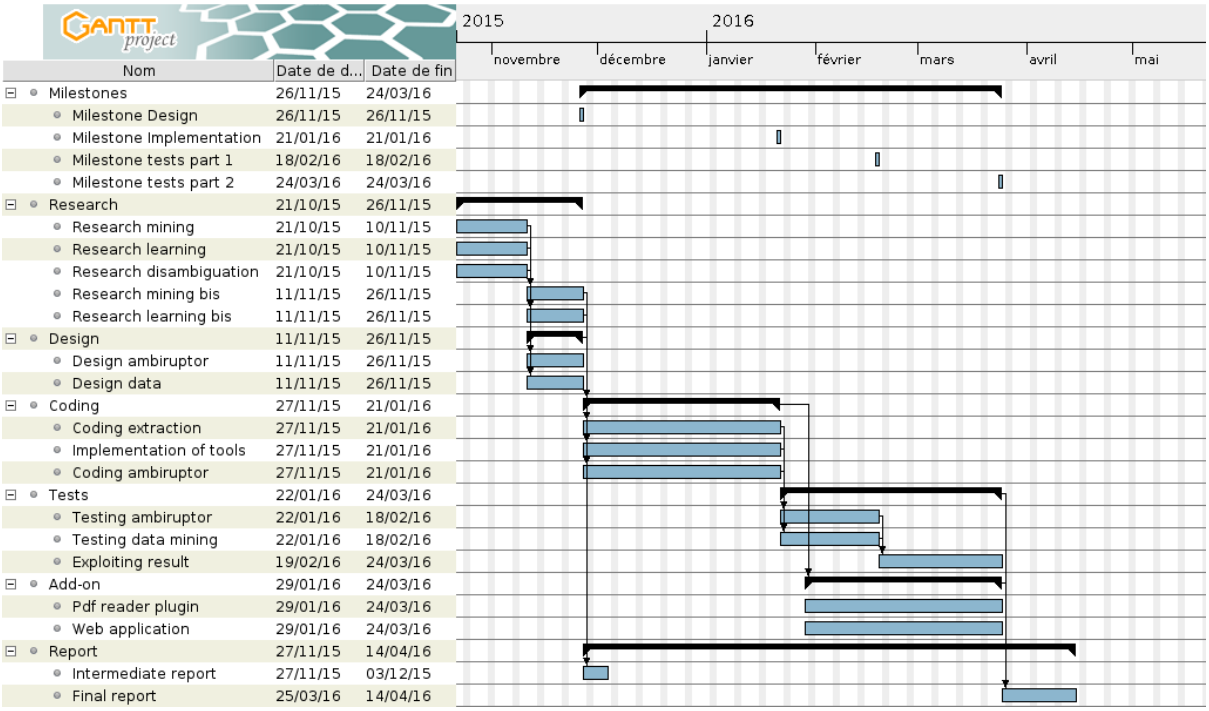
All students will summarize their year's work. Two or three people will compile those documents and write the final report.

3 Organization

3.1 Project Management

- Git repository: `https://github.com/Ambiruptor`
- Mailing list: `ambiruptor@ens-lyon.fr`

3.2 Gantt chart



Nom	Date de d...	Date de fin
Milestones	26/11/15	24/03/16
Milestone Design	26/11/15	26/11/15
Milestone Implementation	21/01/16	21/01/16
Milestone tests part 1	18/02/16	18/02/16
Milestone tests part 2	24/03/16	24/03/16
Research	21/10/15	26/11/15
Research mining	21/10/15	10/11/15
Research learning	21/10/15	10/11/15
Research disambiguation	21/10/15	10/11/15
Research mining bis	11/11/15	26/11/15
Research learning bis	11/11/15	26/11/15
Design	11/11/15	26/11/15
Design ambiruptor	11/11/15	26/11/15
Design data	11/11/15	26/11/15
Coding	27/11/15	21/01/16
Coding extraction	27/11/15	21/01/16
Implementation of tools	27/11/15	21/01/16
Coding ambiruptor	27/11/15	21/01/16
Tests	22/01/16	24/03/16
Testing ambiruptor	22/01/16	18/02/16
Testing data mining	22/01/16	18/02/16
Exploiting result	19/02/16	24/03/16
Add-on	29/01/16	24/03/16
Pdf reader plugin	29/01/16	24/03/16
Web application	29/01/16	24/03/16
Report	27/11/15	14/04/16
Intermediate report	27/11/15	03/12/15
Final report	25/03/16	14/04/16

3.3 Milestones

- 26/11/2015 : The architecture of the project is finished.
- 03/12/2015 : Deadline for the intermediate report.
- 21/01/2015 : The implementation is finished.
- 18/02/2015 : First results of the tests.
- 24/03/2015 : Tests are completed.
- 14/04/2015 : Deadline for the final report.