

Bilingual NLP pipeline

Paul Huygen <paul.huygen@huygen.nl>

17th August 2017
14:44 h.

Abstract

This is a description and documentation of the installation of the Newsreader-pipeline¹. It is an instrument to annotate Dutch or English documents with NLP tags. The documents have to be stored in *Newsreader Annotation Format* (NAF [1]).

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 1.1 | Modules of the pipeline | 3 |
| 1.2 | Reproducibility | 3 |
| 2 | Structure of the pipeline | 6 |
| 2.1 | Expected resources | 6 |
| 3 | Construct the infra-structure | 6 |
| 3.1 | File-structure | 8 |
| 3.2 | Download resources | 11 |
| 3.3 | Java | 12 |
| 3.4 | Maven | 13 |
| 3.5 | Maven | 13 |
| 3.6 | Python | 14 |
| 3.6.1 | Python packages | 17 |
| 3.7 | Perl | 18 |
| 3.8 | Spotlight | 21 |
| 3.8.1 | Install spotlight servers | 21 |
| 3.8.2 | Check/start the Spotlight server | 22 |
| 3.9 | Download materials | 28 |
| 4 | Shared libraries | 28 |
| 4.1 | Autoconf | 28 |
| 4.2 | libxml2 and libxslt | 29 |
| 4.3 | Alpino | 30 |
| 4.4 | Treetagger | 30 |
| 4.5 | Timbl and Ticcutils | 33 |
| 4.6 | Svmlib | 34 |
| 4.7 | The Boost library | 35 |

1. <http://www.newsreader-project.eu/files/2012/12/NWR-D4-2-2.pdf>

| | | |
|----------|---|-----------|
| 5 | Install the modules | 35 |
| 5.1 | Parameters in module-scripts | 37 |
| 5.1.1 | Tokeniser | 37 |
| 5.1.2 | Topic detection tool. | 38 |
| 5.1.3 | Morphosyntactic Parser and Alpino | 38 |
| 5.1.4 | Pos tagger | 38 |
| 5.1.5 | Named entity recognition (NERC) | 39 |
| 5.1.6 | Word-sense disambiguation (WSD) | 39 |
| 5.1.7 | NED | 39 |
| 5.1.8 | Dark-entity relinker | 40 |
| 5.1.9 | Heideltime | 40 |
| 5.1.10 | Ontotagger, Framenet-SRL and nominal events | 40 |
| 5.1.11 | NED-reranker | 41 |
| 5.1.12 | Wikify module | 41 |
| 5.1.13 | UKB | 42 |
| 5.1.14 | IMS-WSD | 42 |
| 5.1.15 | Semantic Role labelling | 42 |
| 5.1.16 | SRL server for English | 43 |
| 5.1.17 | srl-Dutch nominals | 43 |
| 5.1.18 | FBK-time, FBK-temprel, FBK-causalrel | 43 |
| 5.1.19 | Factuality | 44 |
| 5.1.20 | Opinion miner | 44 |
| 5.1.21 | Event coreference | 45 |
| 5.1.22 | Corefgraph | 45 |
| 5.2 | Constituent parser | 45 |
| 6 | Utilities | 46 |
| 6.1 | Language detection | 46 |
| 6.2 | Run-script and test-script | 47 |
| 7 | Miscellaneous | 51 |
| 7.1 | Locate the path to the script itself | 51 |
| 7.2 | Logging | 52 |
| A | How to read and translate this document | 52 |
| A.1 | Read this document | 52 |
| A.2 | Process the document | 53 |
| A.3 | The Makefile for this project. | 54 |
| A.4 | Get Nuweb | 55 |
| A.5 | Pre-processing | 55 |
| A.5.1 | Process ‘dollar’ characters | 56 |
| A.5.2 | Run the M4 pre-processor | 56 |
| A.6 | Typeset this document | 56 |
| A.6.1 | Figures | 56 |
| A.6.2 | Bibliography | 58 |
| A.6.3 | Create a printable/viewable document | 58 |
| A.6.4 | Create HTML files | 61 |
| A.7 | Perform the installation | 64 |
| A.8 | Test whether it works | 65 |
| A.9 | Restore paths after transplantation | 65 |
| B | References | 66 |
| B.1 | Literature | 66 |

| | |
|------------------|-----------|
| C Indexes | 66 |
| C.1 Filenames | 66 |
| C.2 Macro's | 66 |
| C.3 Variables | 68 |

1 Introduction

This document describes the installation of a pipeline that annotates texts in order to extract knowledge. The pipeline has been set up as part of the newsreader ² project. It accepts and produces texts in the NAF (Newsreader Annotation Format) format.

Apart from describing the pipeline set-up, the document actually constructs the pipeline. The pipeline has been installed on a (Ubuntu) Linux computer.

The installation has been parameterised. The locations and names that you read (and that will be used to build the pipeline) have been read from variables in file `inst.m4` in the `nuweb` directory.

The installed pipeline is bi-lingual. It is capable to annotate Dutch and English texts. It recognizes the language from the “lang” attribute of the NAF element of the document. Some of the modules are specific for a single language, other modules support both languages. As a result, there must be two pathways to lead a document through the pipeline, one for English and one for Dutch.

The pipeline is a concatenation of independent software modules, each of which reads a NAF document from standard input and produces another NAF document on standard output.

The aim is, to install the pipeline from open-source modules that can e.g. be obtained from Github. However, that aim is only partially fulfilled. Some of the modules still contain elements that are not open-source or data that are not freely available. Because of lack of time, the current version of the installer installs the English pipeline from a frozen repository of the Newsreader Project.

The NLPP pipeline can be seen as constructed in three parts: 1) The software that is needed to run the pipeline, e.g. compilers and interpreters; 2) the modules themselves and 3) scripts to make the modules operate on a document.

1.1 Modules of the pipeline

Table 2 lists the modules in the pipeline. The column *source* indicates the origin of the module. The modules are obtained in one of the following ways:

1. If possible, the module is directly obtained from an open-source repository like Github.
2. Some modules have not been officially published in a repository. These modules have been packed in a tar-ball that can be obtained by the author. In table 2 this has been indicated as SNAPSHOT.

The modules themselves use other utilities like dependency-taggers and POS taggers. These utilities are listed in table 1.

1.2 Reproducibility

An important goal of this pipeline is, to achieve reproducibility. It means, that at some point in the future the annotation could be re-done on the document and it should produce a result that is identical as the result of the original annotation. In our case, reproducibility involves the following aspects:

- The annotated document ought to contain documentation about the annotation process: What modules have been applied, what was the version of the software of each module, Which resources have been used and what was the version of the resources.

2. <http://www.newsreader-project.eu>

| Module | Version | Section | Source |
|------------------|---------|---------|--------------|
| KafNafParserPy | 1.87 | 3.6.1 | Github |
| Alpino | 21088 | 4.3 | RUG |
| Ticcutils | 0.7 | 4.5 | ILK |
| Timbl | 6.4.6 | 4.5 | ILK |
| Treetagger | 3.2 | 4.4 | Uni. München |
| Spotlight server | 0.7 | 3.8 | Spotlight |

Table 1: List of the utilities to be installed. Column description: **directory:** Name of the subdirectory below *mod* in which it is installed; **Source:** From where the module has been obtained; **script:** Script to be included in a pipeline.

| Module | Source | Resources | Section | Commit | Script | language |
|---------------------------|------------------------|----------------|---------|---------|--------------------|----------|
| Tokenizer | Github | Java | 5.1.1 | 1a69... | tok | en/nl |
| Topic detection | Github | Java | 5.1.2 | b332... | topic | en/nl |
| Morpho-syntactic parser | Github | Python, Alpino | 5.1.3 | 7cfb... | mor | nl |
| POS-tagger | snapshot | | ?? | ... | pos | en |
| Named-entity rec/class | Github | | 5.1.5 | b365... | nerc | en/nl |
| Dark-entity relinker | Github | | 5.1.8 | a534... | nerc | en/nl |
| Constituent parser | snapshot | | ?? | ... | constpars | en |
| Word-sense disamb. nl | Github | | 5.1.6 | 6208... | wsd | nl |
| Word-sense disamb. en | snapshot | | 5.1.14 | ... | ewsd | en |
| Named entity/DBP | snapshot | | 5.1.7 | ... | ned | en/nl |
| NED reranker | snapshot | | 5.1.11 | ... | nedrerscript | en |
| Wikify | snapshot | | 5.1.12 | ... | wikify | en |
| UKB | snapshot | | 5.1.13 | ... | ukb | en |
| Coreference-base | snapshot | | 5.1.22 | ... | coref-graph | en |
| Heideltime | Github | | 5.1.9 | 364c... | heidelttime | nl |
| Onto-tagger | Github | | 5.1.10 | 3177... | onto | nl |
| Semantic Role labeling nl | Github | | 5.1.17 | 0602... | srl | nl |
| Semantic Role labeling en | snapshot | | ?? | ... | eSRL | en |
| Nominal Event ann. | Github | | 5.1.10 | 3177... | nomevent | nl |
| SRL dutch nominals | Github | | 5.1.17 | 1c01... | srl-dutch-nominals | nl |
| Framenet-SRL | Github | | 5.1.10 | 3177... | framesrl | nl |
| FBK-time | snapshot | | ?? | ... | FBK-time | en |
| FBK-temprel | snapshot | | ?? | ... | FBK-temprel | en |
| FBK-causalrel | snapshot | | ?? | ... | FBK-causalrel | en |
| Opinion-miner | Github | | 5.1.20 | 93cd... | opinimin | en/nl |
| Event-coref | Github | | 5.1.21 | 24e8... | evcoref | en/nl |
| Factuality tagger | Github | | 5.1.19 | 58fa... | factuality | en |
| Factuality tagger | Github | | 5.1.19 | a09d... | factuality | nl |

Table 2: List of the modules to be installed. Column description: **directory**: Name of the subdirectory below subdirectory *modules* in which it is installed; **source**: From where the module has been obtained; **commit**: Commit-name or version-tag **script**: Script to be included in a pipeline.

- The source code of the modules as well as resources like data-sets and programming languages should be available from open repository.
- The repositories of the resources should use some versioning system enabling to re-use the version that has been used originally.

A problem in some cases is, that we need to use utilities that are supplied by external parties, and we do not have control about their methods of publication and version management. Examples of such utilities are the compilers for programming languages like Java, Python and parsers like Alpino.

Therefore, we have the following policy to achieve reproducibility:

- Each of the modules writes in the output NAF its own version, and details about the used resources in sufficient detail to enable re-processing.
- It is assumed that when a programming language (e.g. Java, Python) is used, annotation can be reproducible when the major versions coincide.
- A script is constructed that reproducibly builds an environment for the pipeline on some software/hardware platform (e.g. Linux on X64 CPU), using utilities that have been stored in some non-open repository (to preclude copyright-problems).

2 Structure of the pipeline

The finished pipeline consists of:

- A directory that contains for each module an directory with the module in installed form.
- A script that reads an input naf file or plain text file from standard in and produces an annotated NAF file on standard out.
- A script that must be “sourced” in order to find the resources that the modules need to find.

The directory with the modules must be relocatable and immutable. That means that scripts in modules do not have write permissions on the module directory and that they have to find other files on path-descriptions relative to the current path of the script itself.

2.1 Expected resources

In order to run the modules expect the following:

- Instruction `java` invokes Java 1.8;
- Instruction `python` invokes Python 3.6;
- Instruction `Perl` invokes Perl 5;
- Variable `TMPDIR` points to a user-writable directory.

3 Construct the infra-structure

In this section we will generate a script that set up an infra-structure in which the pipeline can be exploited. An attempt is made to make as little as possible presumptions about the services that the host provides.

We need to set up the following:

- Java Version 1.8
- Maven (Gradle?)
- Python version 3.6
- Python packages
- Autoconf
- ...

Let us generate a script to do the work:

```
"../env/bin/make_infrastructure" 7a≡
#!/bin/bash
  < get location of the script (7b DIR ) 52a >
  cd $DIR
  source ../../progenv
  echo make_infrastructure 'date':
  echo ' '
  < next part (7c Initialize ) 7p >
  < init make_infrastructure 8e, ... >
  < next part (7d Java ) 7p >
  < set up Java 13a >
  < next part (7e Maven ) 7p >
  < set up Maven 14b >
  < next part (7f Python ) 7p >
  < set up Python 15b, ... >
  < next part (7g autoconf ) 7p >
  < set up autoconf 29d >
  < next part (7h Perl ) 7p >
  < conditionally install Perl 20a >
  < next part (7i Shared libs ) 7p >
  < install shared libs 29f >
  < next part (7j Alpino ) 7p >
  < install Alpino 30a >
  < next part (7k Spotlight ) 7p >
  < install the Spotlight server 22a, ... >
  < next part (7l Treetagger ) 7p >
  < install the treetagger utility 31a, ... >
  < install the ticcutils utility 34a >
  < install the timbl utility 34b >
  < next part (7m Svmlib ) 7p >
  < install svmlib 35a >
  < next part (7n Boost ) 7p >
  < install boost 35b >

  ◇

  < make scripts executable 7o > ≡
    chmod 775 ../env/bin/make_infrastructure
  ◇
```

Fragment defined by 7o, 8c, 17a, 24g, 36a, 46d, 64c.

Fragment referenced in 64d.

```
< next part 7p > ≡
  echo ' '
  echo make_infrastructure 'date': @1
  echo ' '
  ◇
```

Fragment referenced in 7a.

Let us also make a script that cleans up the infra-structure after the installation.

```

"../env/bin/clean_infrastructure" 8a≡
    #!/bin/bash
    < get location of the script (8b DIR ) 52a >
    cd $DIR
    source ../../progenv
    < init make_infrastructure 8e, ... >
    < clean up after installation 14g >
    ◇

< make scripts executable 8c > ≡
    chmod 775 ../env/bin/clean_infrastructure
    ◇

```

Fragment defined by 7o, 8c, 17a, 24g, 36a, 46d, 64c.
 Fragment referenced in 64d.

Before we begin, we can try whether commands that we need to use actually exist and stop execution otherwise.

```

< test presence of command 8d > ≡
    which @1 >/dev/null
    if
        [ $? -ne 0 ]
    then
        echo "Please install @1"
        exit 4
    fi
    ◇

```

Fragment referenced in 8e.
 Uses: install 65a.

```

< init make_infrastructure 8e > ≡
    < test presence of command (8f git ) 8d >
    < test presence of command (8g tar ) 8d >
    < test presence of command (8h unzip ) 8d >
    < test presence of command (8i tcsh ) 8d >
    < test presence of command (8j hg ) 8d >
    ◇

```

Fragment defined by 8e, 11b.
 Fragment referenced in 7a, 8a.

3.1 File-structure

Let us set up the pipeline in a directory-structure that looks like figure 1. The directories have the following functions.

socket: The directory in the host where the pipeline is to be implemented.

root: The root of the pipeline directory-structure.

nuweb: This directory contains this document and everything to create the pipeline from the open sources of the modules.

modules: Contains subdirectories with the NLP modules that can be applied in the pipeline.

bin: Contains for each of the applicable modules a script that reads NAF input, passes it to the module in the modules directory and produces the output on standard out. Furthermore,

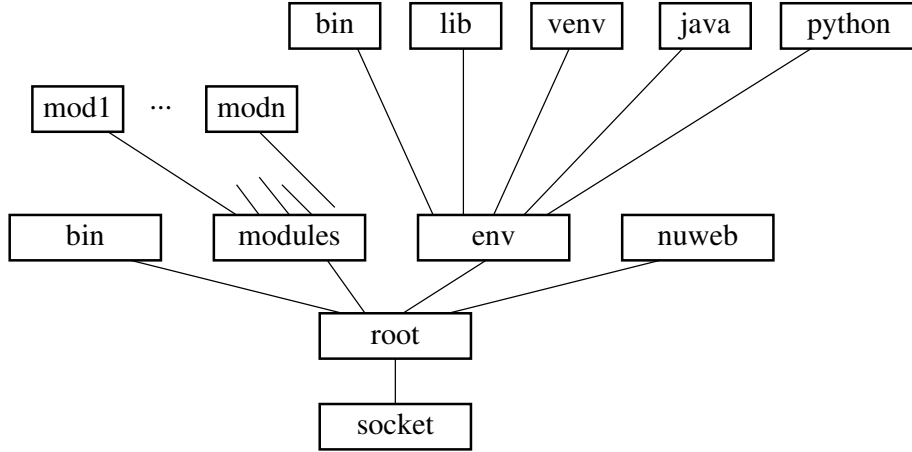


Figure 1: Directory-structure of the pipeline (see text).

the subdirectory contains the script `install_modules` that performs the installation, and a script `test` that shows that the pipeline works in a trivial case.

env: The programming environment. It contains a.o. the Java development kit, Python, the Python virtual environment (**venv**), libraries and binaries.

$\langle \text{directories to create 9a} \rangle \equiv$
`../modules` \diamond

Fragment defined by 9abcd, 59b.
 Fragment referenced in 64b.

$\langle \text{directories to create 9b} \rangle \equiv$
`../bin ../env/bin` \diamond

Fragment defined by 9abcd, 59b.
 Fragment referenced in 64b.

$\langle \text{directories to create 9c} \rangle \equiv$
`../env/lib` \diamond

Fragment defined by 9abcd, 59b.
 Fragment referenced in 64b.

$\langle \text{directories to create 9d} \rangle \equiv$
`../env/etc` \diamond

Fragment defined by 9abcd, 59b.
 Fragment referenced in 64b.

It would be great if an installed pipeline could be moved to another directory while it would keep working. We are not yet sure whether this is possible. However, a minimum condition for this to work would be, that the location of the pipeline can be determined at run-time. To achieve this, let us place a script in the root-directory of the pipeline, that can find in run-time the absolute path to itself and that generates variables that point to the other directories.

```

"../progenv" 10a≡
# Source this script
< get location of the script (10b piperoot ) 52a>
< set variables that point to the directory-structure 10e, ... >
< set environment parameters 10c, ... >
if
  [ -e "$piperoot/progenvv" ]
then
  source $piperoot/progenvv
fi
export progenvset=0
◇

```

Uses: piperoot 10d.

```

< set environment parameters 10c> ≡
export LC_ALL=en_US.UTF-8
export LANG=en_US.UTF-8
export LANGUAGE=en_US.UTF-8
◇

```

Fragment defined by 10c, 30b, 31b, 34e, 35c.

Fragment referenced in 10a.

The full path to the sourced script can be found in variable BASH_SOURCE[0].

```

< find the nlpp root directory 10d> ≡
piperoot="$( cd "$( dirname "${BASH_SOURCE[0]}" )" && pwd )"
◇

```

Fragment never referenced.

Defines: piperoot 10abe, 13e, 16ac, 20c, 29de, 34c, 35a, 36b, 51a.

Once we know piperoot, we know the path to the other directories of figure 1.

```

< set variables that point to the directory-structure 10e> ≡
export pipesocket=${piperoot%*/nlpp}
export nuwebdir=$piperoot/nuweb
export envdir=$piperoot/env
export envbindir=$envdir/bin
export envlibdir=$envdir/lib
export modulesdir=$piperoot/modules
export pipebin=$piperoot/bin
export javadir=$envdir/java
export jarsdir=$javadir/jars
◇

```

Fragment defined by 10ef, 11a, 14f.

Fragment referenced in 10a.

Uses: nuweb 60d, piperoot 10d.

Include a “snapshot” directory that contains non-open materials.

```

< set variables that point to the directory-structure 10f> ≡
export snapshotdir=$pipesocket/v4.0.0.0_nlpp_resources
◇

```

Fragment defined by 10ef, 11a, 14f.

Fragment referenced in 10a.

Add the environment `bin` directory to `PATH`:

```
< set variables that point to the directory-structure 11a > ≡
    export PATH=$envbindir:$pipebin:$PATH
◇
```

Fragment defined by 10ef, 11a, 14f.

Fragment referenced in 10a.

Defines: `PATH` 13e, 14bf, 20c.

3.2 Download resources

To enhance speed of the installation we start to download all resources that we can download at the beginning of the installation in a single blow as parallel processes. We park the resources in a directory `v4.0.0.0_nlpp_resources`, located in the directory where the root of NLPP also resides.

```
< init make_infrastructure 11b > ≡
    < download everything 11c, ... >
    wait
◇
```

Fragment defined by 8e, 11b.

Fragment referenced in 7a, 8a.

Hopefully there will be little to download.

Synchronize with a non-open snapshot-directory if possible. It is only possible if a valid `ssh` key resides in file `nrkey` in the directory in which the `nlpp` root directory resides.

```
< download everything 11c > ≡
    mkdir -p $pipesocket/v4.0.0.0_nlpp_resources
    if
        [ -e /home/paul/nlpp/nrkey ]
    then
        cd $pipesocket
        ( rsync -e "ssh -i /home/paul/nlpp/nrkey" -
          rLt newsreader@kyoto.let.vu.nl:v4.0.0.0_nlpp_resources . ) &
    fi
◇
```

Fragment defined by 11c, 28.

Fragment referenced in 11b.

Download other stuff using `wget`. The following macro downloads a resource into the snapshot-directory if it is not already there.

```
< need to wget 11d > ≡
    if
        [ ! -e $pipesocket/v4.0.0.0_nlpp_resources/@1 ]
    then
        cd $pipesocket/v4.0.0.0_nlpp_resources
        ( wget @2 ) &
    fi
◇
```

Fragment referenced in 13f, 18b, 21b, 29a, 33d.

3.3 Java

We need to have a Java JDK version 1.8 installed. In other words, when we issue the instruction `javac -version` within the pipeline environment, the response must be something like `javac 1.8.0_131`. We assume that if we find a correct Java 1.8, there will also be a proper `java`. Let us first test whether that is the case. If it is not the case, we can install java if a proper tarball is present in the “snapshot directory”.

Let us perform the two tests:

Do we have a proper Java?

```

< check presence of javac in 1.8 12a > ≡
    javac -version 2>&1 | grep 1.8 >/dev/null
    if
        [ $? == 0 ]
    then
        @1="True"
    else
        @1="False"
    fi
    ◇

```

Fragment referenced in 13a.

Do we have a tarball to install Java? (in fact, the following macro can be used to check the presence of any tarball in the snapshot directory).

```

< check whether a tarball is present in the snapshot 12b > ≡
    if
        [ -e $pipesocket/v4.0.0.0_nlpp_resources/@1 ]
    then
        @2="True"
    else
        @2="False"
    fi
    ◇

```

Fragment referenced in 13a, 14b, 15b, 19a.

Now do it:

```

< set up Java 13a > ≡
  < check presence of javac in 1.8 (13b java_OK ) 12a >
  if
    [ ! "$java_OK" == "True" ]
  then
    < check whether a tarball is present in the snapshot (13c jdk-8u131-linux-x64.tar.gz,13d tarball_present ) 12b >
    if
      [ ! "$tarball_present" == "True" ]
    then
      echo "Please install Java 1.8 JDK"
      exit 4
    fi
    mkdir -p $javadir
    cd $javadir
    tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/jdk-8u131-linux-x64.tar.gz
    < set up java environment 13e >
  fi
  ◇

```

Fragment referenced in 7a.

Adapt the PATH variable and set JAVA_HOME. Set these variables in the script that will be sourced in the running pipeline and set them in this script because we are going to need Java.

```

< set up java environment 13e > ≡
  echo 'export JAVA_HOME=$envdir/java/jdk1.8.0_131' >> $piperoot/progenvv
  echo 'export PATH=$JAVA_HOME/bin:$PATH' >> $piperoot/progenvv
  export JAVA_HOME=$envdir/java/jdk1.8.0_131
  export PATH=$JAVA_HOME/bin:$PATH
  ◇

```

Fragment referenced in 13a.

Uses: PATH 11a, piperoot 10d.

3.4 Maven

Currently we need version 3.0.5 to compile the Java sources in some of the modules.

3.5 Maven

Some Java-based modules can best be compiled with [Maven](#). So download and install Maven:

```

< download stuff 13f > ≡
  < need to wget (13g apache-maven-3.0.5-bin.tar.gz,13h http://apache.rediris.es/maven/maven-3/3.0.5/binaries) 13f >
  ◇

```

Fragment defined by 13f, 18b, 21b, 29a, 33d.

Fragment referenced in 28.

First check whether maven is already present in the correct version.

```

< check presence of maven in 3.0.5 14a > ≡
    mvn -version | grep "Maven 3.0.5" >/dev/null
    if
        [ $? == 0 ]
    then
        @1="True"
    else
        @1="False"
    fi
    ◇

```

Fragment referenced in 14b.

```

< set up Maven 14b > ≡
    < check presence of maven in 3.0.5 (14c mvn_OK ) 14a >
    if
        [ ! "$mvn_OK" == "True" ]
    then
        < check whether a tarball is present in the snapshot (14d apache-maven-3.0.5-bin.tar.gz, 14e tarball_present ) 14a >
        if
            [ ! "$tarball_present" == "True" ]
        then
            echo "Please install Maven version 3.0.5"
            exit 4
        fi
        cd $envdir
        tar -xzf /home/paul/nlpp/v4.0.0.0_nlpp_resources/apache-maven-3.0.5-bin.tar.gz
        export MAVEN_HOME=$envdir/apache-maven-3.0.5
        export PATH=${MAVEN_HOME}/bin:${PATH}
    fi
    ◇

```

Fragment referenced in 7a.

```

< set variables that point to the directory-structure 14f > ≡
    export MAVEN_HOME=$envdir/apache-maven-3.0.5
    export PATH=${MAVEN_HOME}/bin:${PATH}
    ◇

```

Fragment defined by 10ef, 11a, 14f.

Fragment referenced in 10a.

Uses: PATH 11a.

When the installation has been finished, we do not need maven anymore.

```

< clean up after installation 14g > ≡
    cd $envdir
    rm -rf apache-maven-3.0.5
    ◇

```

Fragment referenced in 8a.

3.6 Python

Several modules in the pipeline run on Python version 3.6. If the command `python` does not invoke that version, we can try install ActivePython, of which we have a tarball in the snapshot.

Versioning in Python is very confusing. It is the [official Python policy](#) that `/usr/bin/env python` points to Python version 2 but that scripts with a shabang of `#!/usr/bin/env python` should be executable by Python version 2 as well as Python version 3.

Our policy will be as follows:

1. When installing, make sure that command `python3` starts a python 3.6 executable. If this is not the case, install ActivePython version 3.6.
2. Generate a virtual environment.
3. Make sure that in our environmen command `python` executes python from the virtual environment.

```

< check presence of python3 in 3.6 15a > ≡
python3 --version 2>&1 | grep "Python 3.6" >/dev/null
if
[ $? == 0 ]
then
  @1="True"
else
  @1="False"
fi
◇

```

Fragment referenced in [15b](#).

```

< set up Python 15b > ≡
< check presence of python3 in 3.6 (15c python_OK ) 15a >
if
[ ! "$python_OK" == "True" ]
then
  < check whether a tarball is present in the snapshot (15d ActivePython-3.6.0.3600-linux-x86_64-glibc-2.3.6-40)
  if
  [ ! "$tarball_present" == "True" ]
  then
    echo "Please install Python version 3.6"
    exit 4
  fi
  < install ActivePython 16a >
fi
◇

```

Fragment defined by [15b](#), [18a](#).

Fragment referenced in [7a](#).

Unpack the tarball in a temporary directory and install active python in the `env` subdirectory of `nlpp`. Activepython has a few peculiarities:

- It installs things in subdirectories `bin` and `lib` of the installation-directory (in our case subdirectory `env`).
- It installs scripts with names `python3` and `pip3`. We will make symbolic links from these scripts to `python` resp. `pip`.
- It writes self-starting scripts with a “shabang” containing the full absolute path to the `python3` script. In an attempt to make Active-python relocatable we will rewrite the Shabangs to have them contain `#!/usr/bin/env python`.

```

< install ActivePython 16a > ≡
    pytinsdir='mktemp -d -t activepyt.XXXXXX'
    cd $pytinsdir
    tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/ActivePython-3.6.0.3600-linux-x86_64-
    glibc-2.3.6-401834.tar.gz
    accdir='ls -l'
    cd $acdir
    ./install.sh -I $envdir
    cd $piperoot
    rm -rf $pytinsdir
    < create python script and pip script 16b >
    < rewrite ActivePython shabangs 16c >

```

◇

Fragment referenced in 15b.

Uses: install 65a, piperoot 10d.

```

< create python script and pip script 16b > ≡
    cd $envbindir
    rm python
    ln -s python3 python
    rm pip
    ln -s pip3 pip

```

◇

Fragment referenced in 16a.

To rewrite the shabangs of the ActivePython scripts do as follows:

1. Create a temporary directory.
2. Generate an AWK script that replaces the shabang line with a correct one.
3. Generate a script that moves a script from `env/bin` to the temporary directory and then applies the AWK script.
4. Apply the generated script on the scripts in `env/bin`.

```

< rewrite ActivePython shabangs 16c > ≡
    transfile='mktemp -t trans.XXXXXX'
    rm -rf $transfile
    < apply script tran on the scripts in (16d $envbindir, 16e $transfile ) 17c >
    cd $piperoot
    rm -rf $transfile

```

◇

Fragment referenced in 16a.

```

"../env/bin/tran" 16f≡
    #!/bin/bash
    < get location of the script (16g trandir ) 52a >
    workfil=$1
    tempfil=$2
    mv $workfil $tempfil
    gawk -f $strandir/chasbang.awk $tempfil>$workfil
    chmod 775 $workfil

```

◇


```

< make scripts executable 17a > ≡
    chmod 775 ../env/bin/tran
    ◇

```

Fragment defined by 7o, 8c, 17a, 24g, 36a, 46d, 64c.
 Fragment referenced in 64d.

```

"../env/bin/chasbang.awk" 17b ≡
    #!/usr/bin/gawk -f
    BEGIN { shabang="#!/usr/bin/env python3"}

    /\^#\!.*python.*/ { print shabang
                        next
                      }

    {print}
    ◇

```

Uses: `print` 58b.

The following looks complicated. The `find` command applies the `file` command on the files in the `env/bin` directory. The `grep` command filters out the names of the files that are scripts. it produces a filename, followed by a colon, followed by a description of the type of the file. The `gawk` command prints the filenames only and the `xargs` command applies the `tran` script on the file.

```

< apply script tran on the scripts in 17c > ≡
    find @1 -type f -exec file {} + \
    | grep "Python script" | gawk '{print $1}' FS=':' \
    | xargs -iaap $envbindir/tran aap @2
    ◇

```

Fragment referenced in 16c.
 Uses: `print` 58b.

3.6.1 Python packages

In order to be reproducible, we must make sure that Python packages are installed in the correct version. Therefore, we will install the packages beforehand and do not leave that to the install-scripts of the modules. Descriptions of the packages can be found on <https://pypi.python.org>. Install the following packages:

| package | version | module |
|----------------|---------|-----------|
| KafNafParserPy | 1.87 | |
| lxml | 3.8.0 | |
| pyyaml | 3.12 | |
| requests | 2.18.1 | networkx |
| networkx | 1.11 | corefbase |

```

< set up Python 18a > ≡
    pip install KafNafParserPy==1.87
    pip install lxml==3.8.0
    pip install networkx==1.11
    pip install pyyaml==3.12
    pip install requests==2.18.1
    pip install six==1.10.0.
    ◇

```

Fragment defined by 15b, 18a.

Fragment referenced in 7a.

Uses: install 65a.

3.7 Perl

One of the modules uses perl and needs XML::LibXML. However, installation of that package seems to be tricky and seems to depend on the availability of obscure stuff. So, we proceed as follows. First test whether Perl version 5 is present on the host. If that is not the case, check whether we have a tarball named 20160520_nlpp_perllib.tgz in the snapshot. If that is the case, install Perl from scratch and unpack the tarball. Otherwise, fail, and tell the user to install Perl and XML::LibXML.

Install Perl locally, to be certain that Perl is available and to enable to install packages that we need (in any case: XML::LibXML).

```

< download stuff 18b > ≡

    < need to wget (18c perl-5.22.1.tar.gz, 18d http://www.cpan.org/src/5.0/perl-5.22.1.tar.gz ) 11d >

    ◇

```

Fragment defined by 13f, 18b, 21b, 29a, 33d.

Fragment referenced in 28.

```

< check presence of perl in 5 18e > ≡
    perl -v 2>&1 | grep "perl 5," >/dev/null
    if
        [ $? == 0 ]
    then
        @1="True"
    else
        @1="False"
    fi
    ◇

```

Fragment referenced in 19a.

```

< set up Perl 19a > ≡
  < check presence of perl in 5 (19b perl_OK ) 18e >
  if
    [ "$perl_OK" == "True" ]
  then
    < check whether XML::LibXML is installed (19c lib_OK ) 19f >
    if
      [ ! "$lib_OK" == "True" ]
    then
      perl_OK="False"
    fi
  fi
  if
    [ ! "$perl_OK" == "True" ]
  then
    < check whether a tarball is present in the snapshot (19d 20160520_nlpp_perllib.tgz, 19e tarball_present ) 12b >
    if
      [ ! "$tarball_present" == "True" ]
    then
      echo "Please install Perl version 3.6 and XML::LXML"
      exit 4
    fi
    < install perl 20c, ... >
  fi
  fi
  ◇

```

Fragment never referenced.

```

< check whether XML::LibXML is installed 19f > ≡
  perl -MXML::LibXML -e 1 2>/dev/null
  if
    [ $? == 0 ]
  then
    @1="True"
  else
    @1="False"
  fi
  ◇

```

Fragment referenced in 19a.

Install Perl if either Perl is not installed or the installed Perl has a version below 5.

```

< conditionally install Perl 20a > ≡
installPerl=1
which perl >/dev/null
res=$?
if
[ $res -gt 0 ]
then
installPerl=0
else
perlversion='perl -e 'print $];''
if
[ "${perlversion%*.}" -lt 5 ]
then
installPerl=0
fi
fi
if
[ installPerl -gt 0 ]
then
< install Perl 20b >
fi
◇

```

Fragment referenced in 7a.

Uses: `print` 58b.

```

< install Perl 20b > ≡
tmpdir='mktemp -d -t perl.XXXXXX'
cd $tmpdir
tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/perl-5.22.1.tar.gz
cd perl-5.22.1
./Configure -des -Dprefix=$envdir/perl
make
make test
make install
cd $progroot
rm -rf $tmpdir
◇

```

Fragment referenced in 20a.

Uses: `install` 65a.

Make sure that modules use the correct Perl

```

< install perl 20c > ≡
echo 'export PERL_HOME=$envdir/perl' >> $piperoot/progenvv
echo 'export PATH=$PERL_HOME/bin:$PATH' >> $piperoot/progenvv
export PERL_HOME=$envdir/perl
export PATH=$PERL_HOME/bin:$PATH
◇

```

Fragment defined by 20c, 21a.

Fragment referenced in 19a.

Uses: `PATH` 11a, `piperoot` 10d.

Unpack the poor-man tarball with LibXML:

```

< install perl 21a > ≡
    cd $envdir/perl/lib
    tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/20160520_nlpp_perllib.tgz
    ◇

```

Fragment defined by 20c, 21a.
 Fragment referenced in 19a.

3.8 Spotlight

A Spotlight server occupies a lot of memory and we need two of them, one for each language. We may be lucky and have a spotlight server running somewhere. Nevertheless, let us be prepared to be able to install a server ourselves.

3.8.1 Install spotlight servers

Install Spotlight in the way that Itziar Aldabe (<mailto:itziar.aldabe@ehu.es>) described:

The NED module works for English, Spanish, Dutch and Italian. The module returns multiple candidates and correspondences for all the languages. If you want to integrate it in your Dutch or Italian pipeline, you will need:

1. The jar file with the dbpedia-spotlight server. You need the version that Aitor developed in order to correctly use the "candidates" option. You can copy it from the English VM. The jar file name is `dbpedia-spotlight-0.7-jar-with-dependencies-candidates.jar`
2. The Dutch/Italian model for the dbpedia-spotlight. You can download them from: <http://spotlight.sztaki.hu/downloads/>
3. The jar file with the NED module: `ixa-pipe-ned-1.0.jar`. You can copy it from the English VM too.
4. The file: `wikipedia-db.v1.tar.gz`. You can download it from: <http://ixa2.si.ehu.es/ixa-pipes/models/wikipedia-db.v1.tar.gz>. This file contains the required information to do the mappings between the wikipedia-entries. The zip file contains three files: `wikipedia-db`, `wikipedia-db.p` and `wikipedia-db.t`

To start the dbpedia server: Italian server:

```
java -jar -Xmx8g dbpedia-spotlight-0.7-jar-with-dependencies-candidates.jar \
    it http://localhost:2050/rest
```

Dutch server:

```
java -jar -Xmx8g dbpedia-spotlight-0.7-jar-with-dependencies-candidates.jar nl http://localhost:2
```

We set 8Gb for the English server, but the Italian and Dutch Spotlight will require less memory.

So, let us do that.

First, get the Spotlight model data that we need:

```

< download stuff 21b > ≡

    < need to wget (21c nl.tar.gz,21d http://spotlight.sztaki.hu/downloads/archive/2014/nl.tar.gz ) 11d >
    < need to wget (21e en_2+2.tar.gz,21f http://spotlight.sztaki.hu/downloads/archive/2014/en_2+2.tar.gz ) 11d >
    < need to wget (21g wikipedia-db.v1.tar.gz,21h http://ixa2.si.ehu.es/ixa-pipes/models/wikipedia-db.v1.tar.

    ◇

```

Fragment defined by 13f, 18b, 21b, 29a, 33d.
 Fragment referenced in 28.

```

⟨ install the Spotlight server 22a ⟩ ≡
  cd $envdir
  tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/spotlightnl.tgz
  cd $envdir/spotlight
  tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/nl.tar.gz
  tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/en_2+2.tar.gz
  ◇

```

Fragment defined by 22ac.

Fragment referenced in 7a.

```

⟨ get spotlight model ball 22b ⟩ ≡
  if
    [ -e $pipesocket/v4.0.0.0_nlpp_resources/@1 ]
  then
    tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/@1
  else
    wget http://spotlight.sztaki.hu/downloads/archive/2014/@1
    tar -xzf @1
    rm @1
  fi
  ◇

```

Fragment never referenced.

We choose to put the Wikipedia database in the spotlight directory.

```

⟨ install the Spotlight server 22c ⟩ ≡
  cd $envdir/spotlight
  tar -xzf $pipesocket/$snapshotdirectory/wikipedia-db.v1.tar.gz
  ◇

```

Fragment defined by 22ac.

Fragment referenced in 7a.

3.8.2 Check/start the Spotlight server

The macro `check/start spotlight` does the following:

1. Check whether spotlight runs on the default spotlighthost.
2. If that is not the case, and the default host is not `localhost`, check whether Spotlight runs on `localhost`.
3. If a running spotlightserver is still not found, start a spotlightserver on `localhost`.

Start Spotlight, if it doesn't run already. Spotlight ought to run on `localhost` unless variable `spotlighthost` exists. In that case, check whether a Spotlight server can be contacted on that host. Otherwise, change `spotlighthost` to `localhost` and check whether a Spotlight server runs there. If that is not the case, start up a Spotlight server on `localhost`.

The following script, `check_start_spotlight`, has three optional arguments:

language: Default is exported variable `naflang` if it exists, or `en`.

spotlighthost: Name of a host that probably runs a Spotlightserver. Default: exported variable `spotlighthost` if it exists, or `localhost`.

spotlightport: Default: exported variable `spotlightport` if it exists or either 2020 or 2060 for English resp. Dutch.

```
"../bin/check_start_spotlight" 23a≡
    #!/bin/bash
    < get location of the script (23b DIR ) 52a >
    cd $DIR
    source ../progenv
    < get commandline-arguments for check_start_spotlight 23c >
    < set default arguments for Spotlight 24a >
    ◇
```

File defined by 23a, 24b.

The code to obtain command-line arguments has been obtained from [Stackoverflow](#). The following fragment reads the arguments `-l language`, `-h spotlighthost` and `-p spotlightport`:

```
< get commandline-arguments for check_start_spotlight 23c > ≡
    while [[ $# > 1 ]]
    do
        key="$1"

        case $key in
            -l|--language)
                naflang="$2"
                shift # past argument
                ;;
            -h|--spothost)
                spotlighthost="$2"
                shift # past argument
                ;;
            -p|--spotport)
                spotlightport="$2"
                shift # past argument
                ;;
            *)
                # unknown option
                ;;
        esac
        shift # past argument or value
    done
    ◇
```

Fragment referenced in 23a.

Uses: `naflang` 49b.

Fill in default values when they cannot be found in exported variables nor in command-line arguments.

```

< set default arguments for Spotlight 24a > ≡
    if
        [ "$spotlighthost" == "" ]
    then
        spotlighthost=130.37.53.33
    fi
    if
        [ "$spotlightport" == "" ]
    then
        if
            [ "$naflang" == "nl" ]
        then
            spotlightport=2060
        else
            spotlightport=2020
        fi
    fi
    fi
    ◇

```

Fragment referenced in 23a.

Uses: `naflang` 49b.

```

"../bin/check_start_spotlight" 24b≡
    < check listener on host, port (24c $spotlighthost,24d $spotlightport ) 25c >
    if
        [ $spotlightrunning -ne 0 ]
    then
        if
            [ ! "$spotlighthost" == "localhost" ]
        then
            export spotlighthost="localhost"
            < check listener on host, port (24e $spotlighthost,24f $spotlightport ) 25c >
        fi
    fi
    if
        [ $spotlightrunning -ne 0 ]
    then
        < start the Spotlight server on localhost 27a, ... >
    fi
    echo $spotlighthost:$spotlightport
    ◇

```

File defined by 23a, 24b.

```

< make scripts executable 24g > ≡
    chmod 775 ../bin/check_start_spotlight
    ◇

```

Fragment defined by 7o, 8c, 17a, 24g, 36a, 46d, 64c.

Fragment referenced in 64d.

Use function `check_start_spotlight` to find and exploit a running Spotlight-server or to die (with exit code 5) if no server can be found or created. The macro uses implicitly the exported variables `spotlighthost` and `spotlightport` if they exist.


```

⟨ find a spotlightserver or exit 25a ⟩ ≡
    spothostport='/home/paul/nlpp/nlpp/bin/check_start_spotlight -l $naflang'
    export spotlighthost='echo $spothisport | gawk -F ":" '{print $1}''
    export spotlightport='echo $spothisport | gawk -F ":" '{print $2}''
    echo "Spotlight server found on $spothisport." >&2
    if
        [ "$spotlighthost" == "none" ]
    then
        echo "No Spotlight-server found."
        exit 5
    fi
◇

```

Fragment referenced in 51a.

Uses: `naflang` 49b, `print` 58b.

Set the port-number and the language resource for Spotlight, dependent of the language that the user gave as argument.

```

⟨ get spotlight language parameters 25b ⟩ ≡
    if
        [ "$naflang" == "nl" ]
    then
        spotlightport=2060
    else
        spotlightport=2020
    fi
◇

```

Fragment never referenced.

Uses: `naflang` 49b.

The following macro has a hostname and a port-number as arguments. It checks whether something in the host listens on the port and sets variable `success` accordingly:

```

⟨ check listener on host, port 25c ⟩ ≡
    exec 6<>/dev/tcp/01/02 2>/dev/null
    spotlightrunning=$?
    exec 6<&-
    exec 6>&-
◇

```

Fragment referenced in 24b, 27c.

If variable `spotlighthost` does not exist, set it to localhost. Test whether a Spotlightserver runs on `spotlighthost`. If that fails and `spotlighthost` did not point to localhost, try localhost.

If the previous attempts were not succesfull, start the spotlightserver on localhost.

If some spotlightserver has been contacted, set variable `spotlightrunning`. Otherwise exit. At the end variable `spotlighthost` ought to contain the address of the Spotlight-host.

```

< try to obtain a running spotlightserver 26a > ≡
  < test whether spotlighthost runs (26b $spotlighthost ) 26e >
  if
    [ ! $spotlightrunning ]
  then
    if
      [ "$spotlighthost" != "localhost" ]
    then
      export spotlighthost=localhost
      < test whether spotlighthost runs (26c $spotlighthost ) 26e >
    fi
  fi
  if
    [ ! $spotlightrunning ]
  then
    < start the Spotlight server on localhost 27a, ... >
    < test whether spotlighthost runs (26d $spotlighthost ) 26e >
  fi
  if
    [ ! $spotlightrunning ]
  then
    echo "Cannot start spotlight"
    exit 4
  fi
  ◇

```

Fragment never referenced.

Test whether the Spotlightserver runs on a given host. The “spotlight-test” does not really test Spotlight, but it tests whether something is listening on the port and host where we expect Spotlight. I found the test-construction that is used here on [Stackoverflow](#). If the test is positive, set variable `spotlightrunning` to 0. Otherwise, unset that variable.

```

< test whether spotlighthost runs 26e > ≡
  exec 6<>/dev/tcp/01/2060
  if
    [ $? -eq 0 ]
  then
    export spotlightrunning=0
  else
    spotlightrunning=
  fi
  exec 6<&-
  exec 6>&-
  ◇

```

Fragment referenced in 26a.

When trying to start the Spotlight-server on localhost, take care that only one process does this. So we do this:

1. Try to acquire a lock without waiting for it.
2. If we got the lock, run the Spotlight java program in background.
3. If we got the lock, release it.
4. If we did not get the lock, wait for the lock to be released by the process that started the spotlight-server.

But first, we specify the resources for the Spotlight-server.

```

< start the Spotlight server on localhost 27a > ≡
  if
    [ "$naflang" == "nl" ]
  then
    spotresource="nl"
  else
    spotresource="en_2+2"
  fi
  spotlightjar=dbpedia-spotlight-0.7-jar-with-dependencies-candidates.jar
  ◇

```

Fragment defined by 27ab.

Fragment referenced in 24b, 26a.

Uses: naflang 49b.

```

< start the Spotlight server on localhost 27b > ≡
  local oldd='pwd'
  cd /home/paul/nlpp/nlpp/env/spotlight
  $envbindir/sematree acquire spotlock 0
  gotit=$?
  if
    [ $gotit == 0 ]
  then
    java -jar -Xmx8g $spotlightjar $spotresource \
      http://localhost:$spotlightport/rest &
    < wait until the spotlight server is up or faulty 27c >
    $envbindir/sematree release spotlock
  else
    < wait until the spotlight server is up or faulty 27c >
  fi
  cd $oldd
  ◇

```

Fragment defined by 27ab.

Fragment referenced in 24b, 26a.

When the Sportlight server has been started, it takes up to a minute until it really listens on its port. When there is something wrong, it will never listen, of course. Therefore, we give it three minutes. If after that time still nothing listens, we set `spotlighthost` to `none`, indicating that something has gone wrong.

```

< wait until the spotlight server is up or faulty 27c > ≡
  trial=0
  maxtrials=12
  while
    trial=$((trial+1))
    < check listener on host, port (27d $spotlighthost, 27e $spotlightport ) 25c >
    [ $spotlightrunning -ne 0 ] && [ $trial -le $maxtrials ]
  do
    sleep 10
  done
  if
    [ $spotlightrunning -ne 0 ]
  then
    export spotlighthost="none"
  fi
  ◇

```

Fragment referenced in 27b.

Start the Spotlight if it is not already running. First find out what the host is on which we may expect to find a listening Spotlight.

Variable `spotlighthost` contains the address of the host where we expect to find Spotlight. If the expectation does not come true, and the Spotlighthost was not localhost, test whether Spotlight can be found on localhost. If the spotlight-server cannot be found, start it up on localhost.

3.9 Download materials

This installer needs to download a lot from different sources:

- Most of the NLP-modules will be built up from their sources in Github. The sources must be cloned.
- Many modules need external resources, e.g. the Alpino tagger. Often these utilities must be downloaded from a location specified by the supplier.
- Many modules use extra resources like model-data, that must be obtained separately.
- Some of the resources are not publicly available. They must be obtained from a pass-word protected URL.
-

Usually downloads are slow, and the duration is only little determined by the resources in the installing computer, but by the network and the performance of the systems from which we download. Therefore, we may speed up by first downloading things, if possible in parallel processes.

We put the following the beginning of the install-script:

```
< download everything 28 > ≡
  < download stuff 13f, ... >
  echo Waiting for downloads to complete ...
  wait
  echo Download completed
  ◇
```

Fragment defined by 11c, 28.

Fragment referenced in 11b.

4 Shared libraries

When we do not want to rely on what the host can present to us, we need to make our own shared libraries. For the present, we will generate the shared libraries `libxslt` and `libxml2`. We do the following:

1. install autoconf, needed to compile the libs.
2. install libxslt
3. install libxml2

4.1 Autoconf

Gnu autoconf is a system to help configure the Makefiles for a software package. Softwarepackages that use this, supply a file `configure`, `configure.in` or `configure.ac`. To compile and install a package from source we can then perform 1) `./configure --prefix=<environment>`; 2) `make`; 3) `make install`.

Get autoconf:

< download stuff 29a > ≡

< need to wget (29b autoconf-2.69.tar.gz, 29c http://ftp.gnu.org/gnu/autoconf/autoconf-2.69.tar.gz) 11d >
 ◇

Fragment defined by 13f, 18b, 21b, 29a, 33d.

Fragment referenced in 28.

Install autoconf:

< set up autoconf 29d > ≡

```
autoconfdir='mktemp -d -t autoconf.XXXXXX'
cd $autoconfdir
tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/autoconf-2.69.tar.gz
cd autoconf-2.69
./configure --prefix=$envdir
make
make install
cd $piperoot
rm -rf $autoconfdir
```

◇

Fragment referenced in 7a.

Uses: install 65a, piperoot 10d.

4.2 libxml2 and libxslt

Compilation and installation of libxml2 and libxslt goes similar, according to the following template:

< install libxml2 or libxslt 29e > ≡

```
shtmpdir='mktemp -d -t shl.XXXXXX'
cd $shtmpdir
git clone @1
packagedir='ls -1'
cd $packagedir
./autogen.sh --prefix=$envdir
make
make install
cd $piperoot
rm -rf $shtmpdir
```

◇

Fragment referenced in 29f.

Uses: install 65a, piperoot 10d.

< install shared libs 29f > ≡

< install libxml2 or libxslt (29g git://git.gnome.org/libxml2) 29e >
< install libxml2 or libxslt (29h git://git.gnome.org/libxslt) 29e >
 ◇

Fragment referenced in 7a.

4.3 Alpino

Install Alpino as a utility because it is so big, and hard to install on different platforms. Users may choose to install the utilities (and Alpino) by hand and then still install the modules with the script from this file.

Alpino cannot be obtained from an open source repository and there does not seem to be a repository where all the older versions are stored. Therefore, if possible, we will use a copy from our secret archive if that is available. If that is not available, we will download the latest version of Alpino.

```

< install Alpino 30a > ≡
  alpinosrc=Alpino-x86_64-Linux-glibc-2.19-21088-sicstus.tar.gz
  cd $envdir
  if
  [ -d "Alpino" ]
  then
    echo "Not installing Alpino, because of existing directory $envdir/Alpino"
  else
    if
      [ ! -e "$pipesocket/v4.0.0.0_nlpp_resources/$alpinosrc" ]
    then
      echo "Try to install the latest Alpino."
      alpinosrc=latest.tar.gz
      cd $pipesocket/v4.0.0.0_nlpp_resources
      wget http://www.let.rug.nl/vannoord/alp/Alpino/versions/binary/latest.tar.gz
      if
        [ $? -gt 0 ]
      then
        echo "Cannot install Alpino. Please install Alpino in $envdir/Alpino"
        exit 4
      fi
    fi
    cd $envdir
    tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/$alpinosrc
  fi
  ◇

```

Fragment referenced in 7a.

Uses: install 65a.

```

< set environment parameters 30b > ≡
  export ALPINO_HOME=$envdir/Alpino
  ◇

```

Fragment defined by 10c, 30b, 31b, 34e, 35c.

Fragment referenced in 10a.

Defines: ALPINO_HOME Never used.

4.4 Treetagger

Installation of Treetagger goes as follows (See [Treetagger's homepage](#)):

1. Download and unpack the Treetagger tarball. This generates the subdirectories `bin`, `cmd` and `doc`
2. Download and unpack the tagger-scripts tarball

The location where Treetagger comes from and the location where it is going to reside:

```

< install the treetagger utility 31a > ≡
    TREETAGDIR=treetagger
    TREETAGGER_HOME=$envdir/$TREETAGDIR
    TREETAG_BASIS_URL=http://www.cis.uni-muenchen.de/%7Eschmid/tools/TreeTagger/data/
    ◇

```

Fragment defined by 31acde, 32ab, 33bc.

Fragment referenced in 7a.

Defines: TREETAGGER_HOME 31b, 32d, 33a.

```

< set environment parameters 31b > ≡
    export TREETAGGER_HOME=$envdir/treetagger
    ◇

```

Fragment defined by 10c, 30b, 31b, 34e, 35c.

Fragment referenced in 10a.

Uses: TREETAGGER_HOME 31a.

The source tarball, scripts and the installation-script:

```

< install the treetagger utility 31c > ≡
    TREETAGSRC=tree-tagger-linux-3.2.1.tar.gz
    TREETAGSCRIPTS=tagger-scripts.tar.gz
    TREETAG_INSTALLSCRIPT=install-tagger.sh
    ◇

```

Fragment defined by 31acde, 32ab, 33bc.

Fragment referenced in 7a.

Uses: install 65a.

Parametersets:

```

< install the treetagger utility 31d > ≡
    DUTCHPARS_UTF_GZ=dutch-par-linux-3.2-utf8.bin.gz
    DUTCH_TAGSET=dutch-tagset.txt
    DUTCHPARS_2_GZ=dutch2-par-linux-3.2-utf8.bin.gz
    ◇

```

Fragment defined by 31acde, 32ab, 33bc.

Fragment referenced in 7a.

Download everything in the target directory:

```

< install the treetagger utility 31e > ≡
    mkdir -p $envdir/$TREETAGDIR
    cd $envdir/$TREETAGDIR
    wget $TREETAG_BASIS_URL/$TREETAGSRC
    wget $TREETAG_BASIS_URL/$TREETAGSCRIPTS
    wget $TREETAG_BASIS_URL/$TREETAG_INSTALLSCRIPT
    wget $TREETAG_BASIS_URL/$DUTCHPARS_UTF_GZ
    wget $TREETAG_BASIS_URL/$DUTCH_TAGSET
    wget $TREETAG_BASIS_URL/$DUTCHPARS_2_GZ
    ◇

```

Fragment defined by 31acde, 32ab, 33bc.

Fragment referenced in 7a.

Run the install-script:

```

< install the treetagger utility 32a > ≡
    chmod 775 $TREETAG_INSTALLSCRIPT
    ./ $TREETAG_INSTALLSCRIPT
    ◇

```

Fragment defined by 31acde, 32ab, 33bc.
 Fragment referenced in 7a.

The scripts in the `cmd` subdirectory contain absolute paths. We can make the treetagger directory-structure location-independent by using relative paths, eg relative to `TREETAGGER_HOME`

```

< install the treetagger utility 32b > ≡
    < make treetagger location-independent 32c >
    ◇

```

Fragment defined by 31acde, 32ab, 33bc.
 Fragment referenced in 7a.

It works as follows:

Many of the scripts in the `cmd` subdirectory contain lines like:

```
BIN=<absolute path>/bin
```

We read one of those scripts and extract the contents of `<absolute path>` into variable `indicator`. Then we replace in all scripts occurrences of this text with `${TREETAGGER_HOME}`.

```

< make treetagger location-independent 32c > ≡
    < extract the absolute path from one of the scripts 32d >
    < replace the absolute paths 33a >
    ◇

```

Fragment referenced in 32b.

```

< extract the absolute path from one of the scripts 32d > ≡
    cmdir=${TREETAGGER_HOME}/cmd
    probefile='grep -l "^BIN=" ${cmdir}/* | head -n 1'
    indicator='cat $probefile | gawk '/^BIN=/ {< matchscript 32e >}' '
    ◇

```

Fragment referenced in 32c.
 Uses: `TREETAGGER_HOME` 31a.

```

< matchscript 32e > ≡
    match($0, /^BIN=(.*treetagger)\.bin/, arr); print arr[1]◇

```

Fragment referenced in 32d.
 Uses: `print` 58b.


```

< replace the absolute paths 33a > ≡
    sedcommand="s|$indicator|\\${TREETAGGER_HOME}|g"
    tempfile='mktemp -t mytemp.XXXXXX'
    for file in ${cmdir}/*
    do
        mv $file $tempfile
        cat $tempfile | sed $sedcommand >$file
    done
    rm -rf $tempfile
◇

```

Fragment referenced in 32c.

Uses: TREETAGGER_HOME 31a.

Make the treetagger utilities available for everybody.

```

< install the treetagger utility 33b > ≡
    chmod -R o+rx $envdir/$TREETAGDIR/bin
    chmod -R o+rx $envdir/$TREETAGDIR/cmd
    chmod -R o+r $envdir/$TREETAGDIR/doc
    chmod -R o+rx $envdir/$TREETAGDIR/lib
◇

```

Fragment defined by 31acde, 32ab, 33bc.

Fragment referenced in 7a.

Remove the tarballs:

```

< install the treetagger utility 33c > ≡
    rm $TREETAGSRC
    rm $TREETAGSCRIPTS
    rm $TREETAG_INSTALLSCRIPT
    rm $DUTCHPARS_UTF_GZ
    rm $DUTCH_TAGSET
    rm $DUTCHPARS_2_GZ
◇

```

Fragment defined by 31acde, 32ab, 33bc.

Fragment referenced in 7a.

4.5 Timbl and Ticcutils

Timbl and Ticcutils are installed from their source-tarballs. The installation is not (yet?) completely reproducibile because it uses the C-compiler that happens to be available on the host. Installation involves:

1. Download the tarball in a temporary directory.
2. Unpack the tarball.
3. cd to the unpacked directory and perform `./configure`, `make` and `make install`. Note the argument that causes the files to be installed in the `lib` and the `bin` sub-directories of the `env` directory.

```

< download stuff 33d > ≡

```

```

    < need to wget (33e timbl-6.4.6.tar.gz,33f http://software.ticc.uvt.nl/timbl-6.4.6.tar.gz ) 11d >
    < need to wget (33g ticcutils-0.7.tar.gz,33h http://software.ticc.uvt.nl/ticcutils-0.7.tar.gz ) 11d >
◇

```

Fragment defined by 13f, 18b, 21b, 29a, 33d.

Fragment referenced in 28.

```

< install the ticcutils utility 34a > ≡
    URL=http://software.ticc.uvt.nl/ticcutils-0.7.tar.gz
    TARB=ticcutils-0.7.tar.gz
    DIR=ticcutils-0.7
    < unpack ticcutils or timbl 34c >
    ◇

```

Fragment referenced in 7a, 34d.

```

< install the timbl utility 34b > ≡
    TARB=timbl-6.4.6.tar.gz
    DIR=timbl-6.4.6
    < unpack ticcutils or timbl 34c >
    ◇

```

Fragment referenced in 7a, 34d.

```

< unpack ticcutils or timbl 34c > ≡
    SUCCES=0
    ticbeldir='mktemp -t -d tickbel.XXXXXX'
    cd $ticbeldir
    tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/$TARB
    cd $DIR
    sh ./bootstrap.sh
    ./configure --prefix=$envdir
    make
    make install
    cd $piperoot
    rm -rf $ticbeldir
    ◇

```

Fragment referenced in 34ab.

Uses: install 65a, piperoot 10d.

```

< re-install modules after the transplantation 34d > ≡
    < install the ticcutils utility 34a >
    < install the timbl utility 34b >
    ◇

```

Fragment never referenced.

4.6 Svmlib

Svmlib is needed by module svmwsd. That module can install svmlib by itself, but for now we try installation in the prog-environment. We set variable SVMLIB_HOME to indicate where the module is located.

```

< set environment parameters 34e > ≡
    export SVMLIB_HOME=$envdir/svmlib
    ◇

```

Fragment defined by 10c, 30b, 31b, 34e, 35c.

Fragment referenced in 10a.

Defines: SVMLIB_HOME 35a.

```

< install svmllib 35a > ≡
    export SVMLIB_HOME=${envdir}/svmllib
    tempdir='mktemp -d -t svmllib.XXXXXX'
    cd $tempdir
    wget https://github.com/cjlin1/libsvm/archive/master.zip
    unzip master.zip
    rm master.zip
    oridir='ls -1 | head -1'
    mv $oridir ${SVMLIB_HOME}
    cd ${SVMLIB_HOME}/python
    rm -rf $tempdir
    make
    cd $piperoot
    ◇

```

Fragment referenced in 7a.

Uses: piperoot 10d, SVMLIB_HOME 34e.

4.7 The Boost library

I have no idea how Boost works. Neither can I find out how to test whether boost has been installed already. So we install libboost according to [this manual](#) and hope for the best.

```

< install boost 35b > ≡
    cd $envdir
    tar -xzf $pipesocket/v4.0.0.0_nlpp_resources/20160103_boost_1_54_bin.tgz
    ◇

```

Fragment referenced in 7a.

Zet de boost libraries in LD_LIBRARY_PATH.

```

< set environment parameters 35c > ≡
    export LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:${envdir}/boost_1_54_0/stage/lib
    ◇

```

Fragment defined by 10c, 30b, 31b, 34e, 35c.

Fragment referenced in 10a.

5 Install the modules

We make a separate script to install the modules. By default, the modules will be installed in subdirectory `modules` of the NLPP root directory, but this is not necessarily so.

The script `install_modules` installs modules that are not yet present.

```

"../env/bin/install_modules" 35d≡
    #!/bin/bash
    < get location of the script (35e DIR ) 52a >
    cd $DIR
    source ../../progenv
    < variables of the module-installer 52b >
    < functions of the module-installer 36b >
    < install the modules 37d, ... >
    ◇

```

```

< make scripts executable 36a > ≡
    chmod 775 ../env/bin/install_modules
    ◇

```

Fragment defined by 7o, 8c, 17a, 24g, 36a, 46d, 64c.
 Fragment referenced in 64d.

Installing a module from Github is very simple:

- Skip installation if the module is already present. Otherwise:
- Clone the module in subdirectory `modules`.
- `cd` to that module and perform script `install`.

```

< functions of the module-installer 36b > ≡
    function gitinst () {
        url=$1
        dir=$2
        commitset=$3
        echo "Install $dir" >&2
        cd $piperoot/modules
        if
            [ -e $dir ]
        then
            echo "Not installing existing module $dir"
        else
            git clone $url
            cd $dir
            git checkout $commitset
            ./install
        fi
    }
    ◇

```

Fragment referenced in 35d.
 Uses: `install` 65a, `piperoot` 10d.

For each module we generate a script in the `bin` subdirectory to make the module easier to use. The script does the following:

1. Find the directory of itself.
2. Run script `run` in the directory of the module, that can be found as `../<modulename>/run`.

```

< contents of shorthand-script 36c > ≡
    #!/bin/bash
    < get location of the script (36d thisdir ) 52a >
    scriptname=${0##*/}
    scriptpath=$thisdir/$scriptname
    cd ${thisdir}
    < set the naflang parameter 37a >
    cat | ../modules/@1/run
    ◇

```

Fragment referenced in 38adgj, 39beh, 40be, 41bdfi, 42adg, 43ae, 44e, 45be, 46a.

```

⟨ set the naflang parameter 37a ⟩ ≡
    if
        [ -z "${naflang}" ]
    then
        naffile='mkttemp -t naf.XXXXXX'
        cat >$naffile
        naflang='cat $naffile | python $envbindir/langdetect.py'
        export naflang
        cat $naffile | $scriptpath
        result=$?
        rm $naffile
        exit $result
    fi

```

◇

Fragment referenced in 36c, 44b.
 Uses: **naflang** 49b.

5.1 Parameters in module-scripts

Some modules need parameters. All modules need a language specification. The language can be passed as exported variable **naflang**, but it can also be passed as argument **-l**. Furthermore, some modules need contact with a Spotlight server. With the arguments **-h** and **-b** the host and port of a running Spotlight-server can be passed.

Let us assess a “Parameter-passing” hierarchy for **run** scripts. Basically a “run” script uses default values encoded in the **run** script itself. These values can be overruled by environment parameters. Both default and environment parameter settings can be overruled by options that are provided to the **run** commands.

Let us adhere to the policy that we use short one-letter options in **run** scripts, that can be parsed with **getopts**.

The code to obtain command-line arguments in Bash has been obtained from [Stackoverflow](#). The following fragment reads the arguments **-l language**, **-h spotlighthost** and **-p spotlightport**:

```

⟨ start of module-script 37b ⟩ ≡
    ⟨ get location of the script (37c DIR ) 52a ⟩
    cd $DIR
    source ../progenv

```

◇

Fragment never referenced.

5.1.1 Tokeniser

The tokenizer is the simplest of the modules. It needs Java version 1.8. On installation it compiles a Java JAR file, and this is used in the run script.

```

⟨ install the modules 37d ⟩ ≡
    gitinst https://github.com/PaulHuygen/ixa-pipe-tok.git ixa-pipe-
    tok 1a69dbbf337aaf7a97bd21dffcfdbd7cb8ab0d83

```

◇

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```
"../bin/tok" 38a≡
  ⟨ contents of shorthand-script (38b ixa-pipe-tok ) 36c ⟩
  ◇
```

5.1.2 Topic detection tool.

The topic detection tool uses Java.

```
⟨ install the modules 38c ⟩ ≡
  gitinst https://github.com/PaulHuygen/ixa-pipe-topic.git ixa-pipe-
  topic b33259ec587b7ead20d9a2cc72d3c68bdbbae163
  ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
Fragment referenced in 35d.

```
"../bin/topic" 38d≡
  ⟨ contents of shorthand-script (38e ixa-pipe-topic ) 36c ⟩
  ◇
```

5.1.3 Morphosyntactic Parser and Alpino

The morphosyntactic parser is in fact a wrapper around Alpino. We have installed Alpino in section ???. The morpho-syntactic parser expects Alpino to be located in \$envdir/Alpino.

```
⟨ install the modules 38f ⟩ ≡
  gitinst https://github.com/PaulHuygen/morphosyntactic_parser_nl.git morphosyn-
  tactic_parser_nl 7cfb22ed99e9e72966da5dcafe5527628c16d16
  ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
Fragment referenced in 35d.

```
"../bin/mor" 38g≡
  ⟨ contents of shorthand-script (38h morphosyntactic_parser_nl ) 36c ⟩
  ◇
```

5.1.4 Pos tagger

Use the pos-tagger from EHU for English documents.

```
⟨ install the modules 38i ⟩ ≡
  gitinst git@github.com:PaulHuygen/ixa-pipe-pos.git ixa-pipe-
  pos 518fe51d3f196f0ea5695811425128181565b5d7
  ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
Fragment referenced in 35d.

```
"../bin/pos" 38j≡
  ⟨ contents of shorthand-script (38k ixa-pipe-pos ) 36c ⟩
  ◇
```

5.1.5 Named entity recognition (NERC)

```

< install the modules 39a > ≡
    gitinst git@github.com:PaulHuygen/ixa-pipe-nerc.git ixa-pipe-
    nerc b365a180e3e9989f2ff4afcb5957290bc4bfe45f
    ◇

```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```

"../bin/nerc" 39b ≡
    < contents of shorthand-script (39c ixa-pipe-nerc ) 36c >
    ◇

```

5.1.6 Word-sense disambiguation (WSD)

```

< install the modules 39d > ≡
    gitinst ht-
    tps://github.com/PaulHuygen/svm_wsd.git svm_wsd 62080274247e2dd32226e730776f2d447e90e753
    ◇

```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```

"../bin/wsd" 39e ≡
    < contents of shorthand-script (39f svm_wsd ) 36c >
    ◇

```

5.1.7 NED

The NED module is rather picky about the structure of the NAF file. In any case, it does not accept a file that has been produced by the ontotagger. Hence, in a pipeline NED should be executed before the ontotagger.

The NED module wants to consult the Dbpedia Spotlight server, so that one has to be installed somewhere. For this moment, let us suppose that it has been installed on localhost.

```

< install the modules 39g > ≡
    gitinst git@github.com:PaulHuygen/ixa-pipe-ned.git ixa-pipe-
    ned 0917a095e719ce909da5db97a5f36934cbc4f5e5
    ◇

```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```

"../bin/ned" 39h ≡
    < contents of shorthand-script (39i ixa-pipe-ned ) 36c >
    ◇

```

5.1.8 Dark-entity relinker

The “Dark Entity Relinker” tries to link “Dark entities” (named entities that have not been recognized) to the link of a known entity with a similar name structure that has been found in the same text.

```
⟨ install the modules 40a ⟩ ≡
    gitinst git@github.com:PaulHuygen/entity-relink-pipeline.git entity-relink-
    pipeline a534c5708c02217cc9bd1a3369db9c6a056711c0
    ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```
"../bin/derel" 40b ≡
    ⟨ contents of shorthand-script (40c entity-relink-pipeline ) 36c ⟩
    ◇
```

5.1.9 Heideltime

The code for Heideltime can be found in [Github](#). This repo contains an adapted Jar file.

Use Heideltime via a wrapper, `ixa-pipe-time`, obtained from [Github](#).

Although suggested otherwise, Heideltime seems not to use Treetagger. It works

```
⟨ install the modules 40d ⟩ ≡
    gitinst git@github.com:PaulHuygen/ixa-pipe-time.git ixa-pipe-
    time 364c2700c580d67d2ce9cbb8ebffbf3fa960e383
    ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```
"../bin/heideltime" 40e ≡
    ⟨ contents of shorthand-script (40f ixa-pipe-time ) 36c ⟩
    ◇
```

5.1.10 Ontotagger, Framenet-SRL and nominal events

- Een directory voor drie modules.
- Verwacht module `vua-resources` in een parallele directory.

The three modules `ontotagger` (aka “predicatematrix”), `Framenet-SRL` and nominal event detection are based on the same software packages and resources. The three modules need the same jar `ontotagger-1.0-jar-with-dependencies.jar`, they need resources from the `cltl/vua-resources` Github repository and they are going to execute a script that resides in the `scripts` directory of the `cltl/OntoTagger` repository. So, what we have to do is:

1. Install from the `cltl/OntoTagger` repository.
2. Create the jar and put it in an appropriate place.
3. install from the `cltl/vua-resources` repository.
4. generate a script for each of the modules.

In fact, items 2 and 3 are performed by script `install.sh` from the `OntoTagger` repository.


```

< install the modules 41a > ≡
    gitinst git@github.com:PaulHuygen/OntoTagger.git OntoTag-
    ger 3177a4c64cc44aabbbe9cf96d5fa004a1f2afb19
    ◇

```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

The “Ontotagger” script:

```

"../bin/onto" 41b ≡
    < contents of shorthand-script (41c OntoTagger ) 36c >
    ◇

```

The “Nominal Event Coreference” script:

```

"../bin/nomevent" 41d ≡
    < contents of shorthand-script (41e Nominal_Events ) 36c >
    ◇

```

The “Framenet SRL” script:

```

"../bin/framesrl" 41f ≡
    < contents of shorthand-script (41g Framenet_SRL ) 36c >
    ◇

```

5.1.11 NED-reranker

```

< install the modules 41h > ≡
    gitinst git@github.com:PaulHuygen/NWRDomainModel.git NWRDomain-
    Model 95e7ad06f48841dae98cc07c665effe095bae9d3
    ◇

```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```

"../bin/nedrer" 41i ≡
    < contents of shorthand-script (41j NWRDomainModel ) 36c >
    ◇

```

5.1.12 Wikify module

Wikify needs spotlight.

```

< install the modules 41k > ≡
    gitinst git@github.com:PaulHuygen/ixa-pipe-wikify.git ixa-pipe-
    wikify 90a25e13c3a957178b51264277f69b5f258b7447
    ◇

```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```
"../bin/wikify" 42a≡
  ⟨ contents of shorthand-script (42b ixa-pipe-wikify ) 36c ⟩
◇
```

5.1.13 UKB

The UKB WSD module is up to now only available from closed repositories. There exists a repository [ukb](#) in Git, but this does not seem to include the scripts to process NAF. Therefore, we need to have the repo available beforehand.

```
⟨ install the modules 42c ⟩ ≡
  # UKB
  if
    [ -e $snapshotdir/20170712_EHU-ukb.v30.tgz ]
  then
    cd $modulesdir
    tar -xzf $snapshotdir/20170712_EHU-ukb.v30.tgz
  else
    echo "No UKB"
    exit 1
  fi
◇
```

Fragment defined by [37d](#), [38cfi](#), [39adg](#), [40ad](#), [41ahk](#), [42cfi](#), [43cd](#), [44ad](#), [45adg](#).
 Fragment referenced in [35d](#).

```
"../bin/m4_ukbcript" 42d≡
  ⟨ contents of shorthand-script (42e EHU-ukb.v30 ) 36c ⟩
◇
```

5.1.14 IMS-WSD

```
⟨ install the modules 42f ⟩ ≡
  gitinst git@github.com:PaulHuygen/it_makes_sense_WSD.git it_makes_sense_WSD d8b5564a385cf7b69224c18db
◇
```

Fragment defined by [37d](#), [38cfi](#), [39adg](#), [40ad](#), [41ahk](#), [42cfi](#), [43cd](#), [44ad](#), [45adg](#).
 Fragment referenced in [35d](#).

```
"../bin/ewsd" 42g≡
  ⟨ contents of shorthand-script (42h it_makes_sense_WSD ) 36c ⟩ ◇
```

5.1.15 Semantic Role labelling

```
⟨ install the modules 42i ⟩ ≡
  gitinst git@github.com:PaulHuygen/vua-srl-nl.git vua-srl-
  nl 060264b40e7b856a14408bfa2b56c6c036cfb1fe
◇
```

Fragment defined by [37d](#), [38cfi](#), [39adg](#), [40ad](#), [41ahk](#), [42cfi](#), [43cd](#), [44ad](#), [45adg](#).
 Fragment referenced in [35d](#).

```
"../bin/srl" 43a≡
  ⟨ contents of shorthand-script (43b vua-srl-nl ) 36c ⟩
◇
```

5.1.16 SRL server for English

As far as I know, the English SRL for Newsreader, **EHU-srl-server**, is not yet open-source. Therefore, we still have to rely on the v3.0 version from the Newsreader repository.

This module has been set up as client-server application, making it less suitable for this general pipeline-structure. It means that the server ought to have been started before processing documents.

For now, we only implement the client. The client checks whether some process listens on port 5005 and aborts if that is not the case.

```
⟨ install the modules 43c ⟩ ≡
  # eSRL-server
  if
    [ -e $snapshotdir/20170816_EHU-srl-server.tgz ]
  then
    cd $modulesdir
    tar -xzf $snapshotdir/20170816_EHU-srl-server.tgz
  else
    echo "No eSRL"
    exit 1
  fi
◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
Fragment referenced in 35d.

5.1.17 srl-Dutch nominals

```
⟨ install the modules 43d ⟩ ≡
  gitinst git@github.com:PaulHuygen/vua-srl-dutch-nominal-events.git vua-srl-dutch-
  nominal-events 1c01df721a0411049d7ad4b5d3cd41a3ccd3eeb5
◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
Fragment referenced in 35d.

```
"../bin/srl-dutch-nominals" 43e≡
  ⟨ contents of shorthand-script (43f vua-srl-dutch-nominal-events ) 36c ⟩
◇
```

5.1.18 FBK-time, FBK-temprel, FBK-causalrel

The three modules FBK-time, FBK-temprel, FBK-causalrel are, as far as I know, not open-source yet. So, now we need to install from sbapshot.

5.1.19 Factuality

We have module `vua_factuality` to identify event-factuality in English texts and module `multilingual_factuality` to identify event-factuality in non-English texts.

```
< install the modules 44a > ≡
  gitinst git@github.com:PaulHuygen/multilingual_factuality.git multilin-
  gual_factuality a09d815212e047a1cef7109e8c299c9913278d67
  gitinst git@github.com:PaulHuygen/vua_factuality.git vua_factuality a09d815212e047a1cef7109e8c299c991
  ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.

Fragment referenced in 35d.

The shorthandscrip runs the module in `vua_factuality` for english documents and it runs the module in `multilingual_factuality` for documents in other languages.

```
"../bin/factuality" 44b≡
  #!/bin/bash
  < get location of the script (44c thisdir ) 52a >
  scriptname=${0##*/}
  scriptpath=$thisdir/$scriptname
  < set the naflang parameter 37a >
  cd ${thisdir}
  if
    [ "${naflang}" == "en" ]
  then
    cat | ../modules/vua_factuality/run
  else
    cat | ../modules/multilingual_factuality/run
  fi
  ◇
```

5.1.20 Opinion miner

The opinion-miner needs models that are not yet available from an open repository. The installer expects the variable `opinion_models_ball_path` to contain the full path to the tarball with the opinion-models.

```
< install the modules 44d > ≡
  export opinion_models_ball_path=${snapshotdir}/models_opinion_miner_deluxePP.tgz
  gitinst git@github.com:PaulHuygen/opinion_miner_deluxePP.git opin-
  ion_miner_deluxePP 93cd9e4a35f5964edf8f121b4d18bcb9534bb196
  ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.

Fragment referenced in 35d.

Defines: `opinion_models_ball_path` Never used.

```
"../bin/opinimin" 44e≡
  < contents of shorthand-script (44f opinion_miner_deluxePP ) 36c >
  ◇
```

5.1.21 Event coreference

The event-coreference module is language-independent. It is a module in a jar-file that can be built with the Github [git@github.com:PaulHuygen/EventCoreference.git](https://github.com:PaulHuygen/EventCoreference.git) repo. The module uses resources from the `vua-resources` Github repo.

```
< install the modules 45a > ≡
  gitinst git@github.com:PaulHuygen/EventCoreference.git EventCorefer-
  ence 24e818a23e9e492e24e2b26af8a09baf5d1d6289
  ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```
"../bin/evcoref" 45b≡
  < contents of shorthand-script (45c EventCoreference ) 36c >
  ◇
```

5.1.22 Corefgraph

The corefgraph module is currently still a hacked version of the module that can be found in the newsreader vs. 3.0 repository. It is stored in the snapshot-directory.

So, install the module from there.

```
< install the modules 45d > ≡
  # EHU-corefgraph
  if
    [ -e $snapshotdir/20170814_EHU-corefgraph.v30.tgz ]
  then
    cd $modulesdir
    tar -xzf $snapshotdir/20170712_EHU-ukb.v30.tgz
  else
    echo "No coreference-graph"
    exit 1
  fi
  ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```
"../bin/coref-graph" 45e≡
  < contents of shorthand-script (45f EHU-corefgraph.v30 ) 36c >
  ◇
```

5.2 Constituent parser

```
< install the modules 45g > ≡
  gitinst git@github.com:PaulHuygen/ixa-pipe-parse.git ixa-pipe-
  parse bf8e91f829e963fe16684a12ad241f8b1aab251d
  ◇
```

Fragment defined by 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg.
 Fragment referenced in 35d.

```
"../bin/constpars" 46a≡
  ⟨ contents of shorthand-script (46b ixa-pipe-parse ) 36c ⟩
  ◇
```

6 Utilities

6.1 Language detection

The following script `../env/bin/langdetect.py` discerns the language of the NAF document that it reads from standard in. If it cannot find the language, it prints `unknown`. The macro `set the language variable` uses this script to set variable `naflang`. All pipeline modules expect that this variable has been set.

```
"../env/bin/langdetect.py" 46c≡
  #!/usr/bin/env python
  # langdetect -- Detect the language of a NAF document.
  #
  import xml.etree.ElementTree as ET
  import sys
  import re
  xmldoc = sys.stdin.read()
  #print xmldoc
  root = ET.fromstring(xmldoc)
  # print root.attrib['lang']
  lang = "unknown"
  for k in root.attrib:
    if re.match(".*lang$", k):
      language = root.attrib[k]
  print(language)
  ◇
```

Uses: `print` 58b.

```
⟨ make scripts executable 46d ⟩ ≡
  chmod 775 ../env/bin/langdetect.py
  ◇
```

Fragment defined by 7o, 8c, 17a, 24g, 36a, 46d, 64c.
 Fragment referenced in 64d.

The module-scripts depend on the existence of variable `naflang`. In most cases this is not a problem because the scripts run in a surrounding script that sets `naflang`. However, a users may occasionally run a module-script stand-alone e.g. to debug. In that case, we can read the language from the NAF, set variable `naflang`, and then run the module-script in a subshell. We assume that variable `scriptpath` contains the path of the script itself.

The macro does the following if `naflang` has not been set:

1. Save the content of standard in to a temporary file.
2. Run `langdetect` with the temporary file as input and set the `naflang` variable.
3. Run the script `$scriptpath` (i.e. itself) with the temporary file as input.
4. Remove the temporary file.
5. Exit itself with the errorcode of the sub-script that it has run.

```

⟨run in subshell when naflang is not known 47a⟩ ≡
    if
        [ -z "${naflang+x}" ]
    then
        naffile='mktemp -t naf.XXXXXX'
        cat >$naffile
        naflang='cat $naffile | python $envbindir/langdetect.py'
        export naflang
        cat $naffile | $scriptpath
        result=$?
        rm $naffile
        exit $result
    fi
    ◇

```

Fragment never referenced.

Uses: **naflang** 49b.

```

⟨run only if language is English or Dutch 47b⟩ ≡
    if
        [ ! "$naflang" == "nl" ] && [ ! "$naflang" == "en" ]
    then
        exit 6
    fi
    ◇

```

Fragment never referenced.

Uses: **naflang** 49b.

6.2 Run-script and test-script

The script **nlpp** reads a NAF document from standard in and produces an annotated NAF on standard out. The script **test** annotates either a test-document that resides in the nuweb directory or a user-provided document and leaves the intermediate results in its working directory **nlpp/test**, so that, in case of problems, it is easy traceable what went wrong.

The annotation process involves a sequence in which an NLP module reads a file that contains the output from a previous module (or the input NAF file), processes it and writes the result in another file.

The following function, **runmodule**, performs the action of a single module in the sequence. It needs three arguments: 1) the name of the NAF file that the previous module produced or the input file; 2) the name of directory in which the module resides and 3) the name of the output NAF.

The function uses variable **moduleresult** to decide whether it is really going to annotate. If this variable is "false" (i.e., not equal to zero), this means that one of the previous modules failed, and it is of no use to process the input file. In that case, the function leaves **moderesult** as it is and does not process the input-file. Otherwise, it will process the input-file and it sets **moduleresult** to the result of the processing module.

```

⟨function to run a module 48a⟩ ≡
    export moduleresult=0

    function runmodule {
        local infile=$1
        local modulecommand=$modulesdir/$2/run
        local outfile=$3
        if
            [ $moduleresult -eq 0 ]
        then
            cat $infile | $modulecommand > $outfile
            moduleresult=$?
            if
                [ $moduleresult -gt 0 ]
            then
                failmodule=$modulecommand
                echo "Failed: module $modulecommand; result $moduleresult" >&2
                exit $moduleresult
            else
                echo "Completed: module $modulecommand; result $moduleresult" >&2
            fi
        fi
    }

```

◇

Fragment referenced in 51ab.

Defines: `moduleresult` 51ab, `runmodule` 48b, 49a.

Use the function to annotate a NAF file that `infile` points to and write the result in a file that `outfile` points to:

```

⟨annotate dutch document 48b⟩ ≡
    runmodule $infile      ixa-pipe-tok      tok.naf
    runmodule tok.naf      ixa-pipe-topic    top.naf
    runmodule top.naf      morphosyntactic_parser_nl    pos.naf
    runmodule pos.naf      ixa-pipe-nerc      nerc.naf
    runmodule nerc.naf     svm_wsd            wsd.naf
    runmodule wsd.naf      ixa-pipe-ned       ned.naf
    runmodule ned.naf      entity-relink-pipeline    derel.naf
    runmodule derel.naf    ixa-pipe-time      times.naf
    runmodule times.naf    OntoTagger          onto.naf
    runmodule onto.naf     vua-srl-nl         srl.naf
    runmodule srl.naf      Nominal_Events     nomev.naf
    runmodule nomev.naf    vua-srl-dutch-nominal-events    psrl.naf
    runmodule psrl.naf     Framenet_SRL       fsrl.naf
    runmodule fsrl.naf     multilingual_factuality fact.naf
    runmodule fact.naf     EventCoreference   $outfile

```

◇

Fragment referenced in 49b.

Uses: `runmodule` 48a.

Similar for an English naf:


```

⟨ annotate english document 49a ⟩ ≡
    runmodule $infile      ixa-pipe-tok tok.naf
    runmodule tok.naf      ixa-pipe-topic top.naf
    runmodule top.naf      ixa-pipe-pos pos.naf
    runmodule pos.naf      ixa-pipe-parse parse.naf
    runmodule parse.naf    ixa-pipe-nerc nerc.naf
    runmodule nerc.naf     svm_wsd wsd.naf
    runmodule wsd.naf      ixa-pipe-ned ned.naf
    runmodule ned.naf      entity-relink-pipeline derel.naf
    runmodule derel.naf    NWRDomainModel nedr.naf
    runmodule nedr.naf     ixa-pipe-wikify wikif.naf
    runmodule wikif.naf    EHU-ukb.v30 ukb.naf
    runmodule ukb.naf      it_makes_sense_WSD wsd.naf
    runmodule wsd.naf      EHU-corefgraph.v30 corefg.naf
    runmodule corefg.naf   EHU-srl-server esrl.naf
    runmodule esrl.naf     FBK-time.v30 ftime.naf
    runmodule ftime.naf    FBK-temprel.v30 ftemp.naf
    runmodule ftemp.naf    FBK-causalrel.v30 fcausal.naf
    runmodule fcausal.naf  EventCoreference evcoref.naf
    runmodule evcoref.naf  vua_factuality fact.naf
    ◇

```

Fragment referenced in 49b.

Uses: `runmodule` 48a.

Determine the language and select one of the above macro's to annotate the document. In fact, consider the document as an English document unless `naflang` is "nl"

```

⟨ annotate 49b ⟩ ≡
    naflang='cat $infile | /home/paul/nlpp/nlpp/env/bin/langdetect.py'
    export naflang
    if
        [ "$naflang" == "nl" ]
    then
        ⟨ annotate dutch document 48b ⟩
    else
        ⟨ annotate english document 49a ⟩
    fi
    ◇

```

Fragment referenced in 51ab.

Defines: `naflang` 23c, 24a, 25ab, 27a, 37a, 44b, 47ab, 50.

Use the above "annotate" macro in a test script and in a run script. The scripts set a working directory and put the input-file in it, and then annotate it.

The test-script uses a special test-directory and leaves it behind when it is finished. If the user specified a language, the script copies a NAF testfile from the nuweb directory as input-file. Otherwise, the script expects the test-directory to be present, with an input-file (named `in.naf`) in it.

< get a testfile and set naflang or die 50 > ≡

```

cd $workdir
naflang=""
if
  [ "$1" == "en" ]
then
  cp $nuwebdir/test.en.in.naf $infile
  export naflang="en"
else
  if
    [ "$1" == "nl" ]
  then
    cp $nuwebdir/test.nl.in.naf $infile
    export naflang="nl"
  fi
fi
if
  [ -e $infile ]
then
  if
    [ "$naflang" == "" ]
  then
    naflang='cat $infile | python $envbindir/langdetect.py'
  fi
else
  echo "Please supply test-file $workdir/$infile or specify language"
  exit 4
fi
◇

```

Fragment referenced in [51a](#).

Uses: `naflang` [49b](#).

This is the test-script:

```

"../bin/test" 51a≡
#!/bin/bash
DIR="$( cd "$( dirname "${BASH_SOURCE[0]}" )" && pwd )"
rdir=$(dirname "$DIR")
source $rdir/progenv
oldd='pwd'
workdir=$piperoot/test
mkdir -p $workdir
cd $workdir
infile=in.naf
outfile=out.naf
< get a testfile and set naflang or die 50 >
< find a spotlightserver or exit 25a >
< function to run a module 48a >
< annotate 49b >
if
[ $moduleresult -eq 0 ]
then
echo Test succeeded.
else
echo Something went wrong.
fi
exit $moduleresult
◇

```

Uses: moduleresult 48a, piperoot 10d.

The run-script `nlpp` reads a “raw” naf from standard in and produces an annotated naf on standard out. It creates a temporary directory to store intermediate results from the modules and removes this directory afterwards.

```

"../bin/nlpp" 51b≡
#!/bin/bash
oldd='pwd'
workdir='mktemp -d -t nlpp.XXXXXX'
cd $workdir
cat >$workdir/$infile
< function to run a module 48a >
< annotate 49b >
if
[ $moduleresult -eq 0 ]
then
cat $outfile
fi
cd $oldd
rm -rf $workdir
exit $moduleresult
◇

```

Uses: moduleresult 48a.

7 Miscellaneous

7.1 Locate the path to the script itself

The following macro finds the directory in which the script itself or the sourced script itself is located.

```

⟨ get location of the script 52a ⟩ ≡
    @1="$( cd "$( dirname "${BASH_SOURCE[0]}" )" && pwd )"
    ◇

```

Fragment referenced in 7a, 8a, 10a, 16f, 23a, 35d, 36c, 37b, 44b.

7.2 Logging

Write log messages to standard out if variable LOGLEVEL is equal to 1.

```

⟨ variables of the module-installer 52b ⟩ ≡
    LOGLEVEL=1
    ◇

```

Fragment referenced in 35d.

```

⟨ logmess 52c ⟩ ≡
    if
    [ $LOGLEVEL -gt 0 ]
    then
    echo @1
    fi
    ◇

```

Fragment never referenced.

A How to read and translate this document

This document is an example of *literate programming* [2]. It contains the code of all sorts of scripts and programs, combined with explaining texts. In this document the literate programming tool **nuweb** is used, that is currently available from Sourceforge (URL:nuweb.sourceforge.net). The advantages of Nuweb are, that it can be used for every programming language and scripting language, that it can contain multiple program sources and that it is very simple.

A.1 Read this document

The document contains *code scraps* that are collected into output files. An output file (e.g. `output.fil`) shows up in the text as follows:

```

"output.fil" 4a ≡
    # output.fil
    < a macro 4b >
    < another macro 4c >
    ◇

```

The above construction contains text for the file. It is labelled with a code (in this case 4a) The constructions between the < and > brackets are macro's, placeholders for texts that can be found in other places of the document. The test for a macro is found in constructions that look like:

```

< a macro 4b > ≡
    This is a scrap of code inside the macro.
    It is concatenated with other scraps inside the
    macro. The concatenated scraps replace
    the invocation of the macro.

```

Macro defined by 4b, 87e

Macro referenced in 4a

Macro's can be defined on different places. They can contain other macro's.

< a scrap 87e > \equiv

This is another scrap in the macro. It is concatenated to the text of scrap 4b.

This scrap contains another macro:

< another macro 45b >

Macro defined by 4b, 87e

Macro referenced in 4a

A.2 Process the document

The raw document is named `a_nlpp.w`. Figure 2 shows pathways to translate it into print-

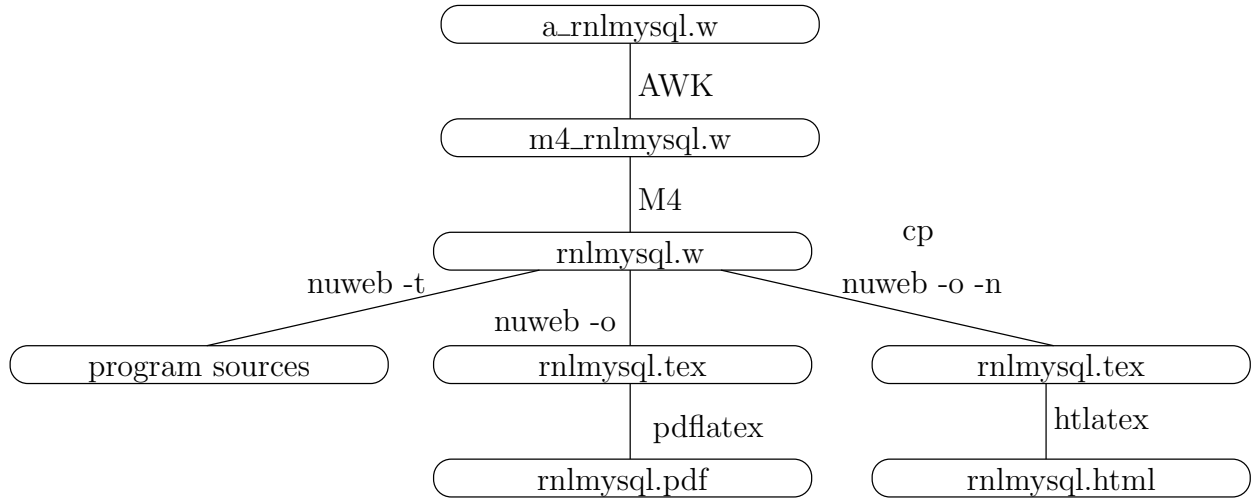


Figure 2: Translation of the raw code of this document into printable/viewable documents and into program sources. The figure shows the pathways and the main files involved.

able/viewable documents and to extract the program sources. Table 3 lists the tools that are

| Tool | Source | Description |
|--------|--|--|
| gawk | www.gnu.org/software/gawk/ | text-processing scripting language |
| M4 | www.gnu.org/software/m4/ | Gnu macro processor |
| nuweb | nuweb.sourceforge.net | Literate programming tool |
| tex | www.ctan.org | Typesetting system |
| tex4ht | www.ctan.org | Convert \TeX documents into <code>xml/html</code> |

Table 3: Tools to translate this document into readable code and to extract the program sources

needed for a translation. Most of the tools (except Nuweb) are available on a well-equipped Linux system.

\langle *parameters in Makefile 54a* $\rangle \equiv$
 NUWEB=../env/bin/nuweb
 ◇

Fragment defined by 54a, 55a, 57ab, 59c, 61b, 64a.
 Fragment referenced in 54b.
 Uses: nuweb 60d.

A.3 The Makefile for this project.

This chapter assembles the Makefile for this project.

"Makefile" 54b \equiv
 \langle *default target 54c* \rangle

 \langle *parameters in Makefile 54a, ...* \rangle

 \langle *impliciete make regels 57c, ...* \rangle
 \langle *expliciete make regels 55b, ...* \rangle
 \langle *make targets 54d, ...* \rangle
 ◇

The default target of make is all.

\langle *default target 54c* $\rangle \equiv$
 all : \langle *all targets 54e* \rangle
 .PHONY : all
 ◇

Fragment referenced in 54b.
 Defines: all Never used, PHONY 58a.

\langle *make targets 54d* $\rangle \equiv$
 clean:
 ../env/bin/clean_infrastructure
 ◇

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.
 Fragment referenced in 54b.

The default is, to install nlpp.

\langle *all targets 54e* $\rangle \equiv$
 install◇

Fragment referenced in 54c.
 Uses: install 65a.

We use many suffixes that were not known by the C-programmers who constructed the `make` utility. Add these suffixes to the list.

< parameters in Makefile 55a > \equiv
`.SUFFIXES: .pdf .w .tex .html .aux .log .php`

◇

Fragment defined by 54a, 55a, 57ab, 59c, 61b, 64a.
 Fragment referenced in 54b.
 Defines: SUFFIXES Never used.
 Uses: pdf 58b.

A.4 Get Nuweb

An annoying problem is, that this program uses nuweb, a utility that is seldom installed on a computer. Therefore, we are going to install that first if it is not present. Unfortunately, nuweb is hosted on sourceforge and it is difficult to achieve automatic downloading from that repository. Therefore I copied one of the versions on a location from where it can be downloaded with a script.

Put the nuweb binary in the nuweb subdirectory, so that it can be used before the directory-structure has been generated.

< expliciete make regels 55b > \equiv

```
nuweb: $(NUWEB)

$(NUWEB): ../nuweb-1.58
    mkdir -p ../env/bin
    cd ../nuweb-1.58 && make nuweb
    cp ../nuweb-1.58/nuweb $(NUWEB)
```

◇

Fragment defined by 55bd, 56ab, 58a, 59d, 61c, 62b.
 Fragment referenced in 54b.
 Uses: nuweb 60d.

< clean up 55c > \equiv
`rm -rf ../nuweb-1.58`

◇

Fragment never referenced.
 Uses: nuweb 60d.

< expliciete make regels 55d > \equiv
`../nuweb-1.58:`
`cd .. && wget http://kyoto.let.vu.nl/~huygen/nuweb-1.58.tgz`
`cd .. && tar -xzf nuweb-1.58.tgz`

◇

Fragment defined by 55bd, 56ab, 58a, 59d, 61c, 62b.
 Fragment referenced in 54b.
 Uses: nuweb 60d.

A.5 Pre-processing

To make usable things from the raw input `a_nlpp.w`, do the following:

1. Process $\$$ characters.
2. Run the m4 pre-processor.
3. Run nuweb.

This results in a \LaTeX file, that can be converted into a PDF or a HTML document, and in the program sources and scripts.

A.5.1 Process ‘dollar’ characters

Many “intelligent” \TeX editors (e.g. the auctex utility of Emacs) handle $\$$ characters as special, to switch into mathematics mode. This is irritating in program texts, that often contain $\$$ characters as well. Therefore, we make a stub, that translates the two-character sequence $\backslash\$$ into the single $\$$ character.

```
<expliciete make regels 56a> ≡
m4_nlpp.w : a_nlpp.w
          gawk '{if(match($$0, "@%")) {printf("%s", substr($$0,1,RSTART-
1))} else print}' a_nlpp.w \
          | gawk '{gsub(/[\$]/, "$$");print}' > m4_nlpp.w
```

◇

Fragment defined by 55bd, 56ab, 58a, 59d, 61c, 62b.

Fragment referenced in 54b.

Uses: print 58b.

A.5.2 Run the M4 pre-processor

```
<expliciete make regels 56b> ≡
nlpp.w : m4_nlpp.w inst.m4
        m4 -P m4_nlpp.w > nlpp.w
```

◇

Fragment defined by 55bd, 56ab, 58a, 59d, 61c, 62b.

Fragment referenced in 54b.

A.6 Typeset this document

Enable the following:

1. Create a PDF document.
2. Print the typeset document.
3. View the typeset document with a viewer.
4. Create a HTMLdocument.

In the three items, a typeset PDF document is required or it is the requirement itself.

A.6.1 Figures

This document contains figures that have been made by `xfig`. Post-process the figures to enable inclusion in this document.

The list of figures to be included:

< parameters in Makefile 57a > ≡
 FIGFILES=fileschema directorystructure

◇

Fragment defined by 54a, 55a, 57ab, 59c, 61b, 64a.
 Fragment referenced in 54b.
 Defines: FIGFILES 57b.

We use the package `figlatex` to include the pictures. This package expects two files with extensions `.pdftex` and `.pdftex_t` for `pdflatex` and two files with extensions `.pstex` and `.pstex_t` for the `latex/dvips` combination. Probably `tex4ht` uses the latter two formats too.

Make lists of the graphical files that have to be present for `latex/pdflatex`:

< parameters in Makefile 57b > ≡
 FIGFILENAMES=\$(foreach fil,\$(FIGFILES), \$(fil).fig)
 PDFT_NAMES=\$(foreach fil,\$(FIGFILES), \$(fil).pdftex_t)
 PDF_FIG_NAMES=\$(foreach fil,\$(FIGFILES), \$(fil).pdftex)
 PST_NAMES=\$(foreach fil,\$(FIGFILES), \$(fil).pstex_t)
 PS_FIG_NAMES=\$(foreach fil,\$(FIGFILES), \$(fil).pstex)

◇

Fragment defined by 54a, 55a, 57ab, 59c, 61b, 64a.
 Fragment referenced in 54b.
 Defines: FIGFILENAMES Never used, PDFT_NAMES 59a, PDF_FIG_NAMES 59a, PST_NAMES Never used,
 PS_FIG_NAMES Never used.
 Uses: FIGFILES 57a.

Create the graph files with program `fig2dev`:

< impliciete make regels 57c > ≡
 %.eps: %.fig
 fig2dev -L eps \$< > \$@

 %.pstex: %.fig
 fig2dev -L pstex \$< > \$@

 .PRECIOUS : %.pstex
 %.pstex_t: %.fig %.pstex
 fig2dev -L pstex_t -p \$*.pstex \$< > \$@

 %.pdftex: %.fig
 fig2dev -L pdftex \$< > \$@

 .PRECIOUS : %.pdftex
 %.pdftex_t: %.fig %.pstex
 fig2dev -L pdftex_t -p \$*.pdftex \$< > \$@

◇

Fragment defined by 57c, 62a.
 Fragment referenced in 54b.
 Defines: `fig2dev` Never used.

A.6.2 Bibliography

To keep this document portable, create a portable bibliography file. It works as follows: This document refers in the `|bibliography|` statement to the local `bib`-file `nlpp.bib`. To create this file, copy the auxiliary file to another file `auxfil.aux`, but replace the argument of the command `\bibdata{nlpp}` to the names of the bibliography files that contain the actual references (they should exist on the computer on which you try this). This procedure should only be performed on the computer of the author. Therefore, it is dependent of a binary file on his computer.

```
< expliciete make regels 58a > ≡
    bibfile : nlpp.aux /home/paul/bin/mkportbib
              /home/paul/bin/mkportbib nlpp litprog

    .PHONY : bibfile
◇
```

Fragment defined by 55bd, 56ab, 58a, 59d, 61c, 62b.

Fragment referenced in 54b.

Uses: PHONY 54c.

A.6.3 Create a printable/viewable document

Make a PDF document for printing and viewing.

```
< make targets 58b > ≡
    pdf : nlpp.pdf

    print : nlpp.pdf
            lpr nlpp.pdf

    view : nlpp.pdf
            evince nlpp.pdf

◇
```

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.

Fragment referenced in 54b.

Defines: pdf 55a, 59a, print 17bc, 20a, 25a, 32e, 46c, 56a, view Never used.

Create the PDF document. This may involve multiple runs of `nuweb`, the `LATEX` processor and the `bibTEX` processor, and depends on the state of the `aux` file that the `LATEX` processor creates as a by-product. Therefore, this is performed in a separate script, `w2pdf`.

The w2pdf script The three processors `nuweb`, `LATEX` and `bibTEX` are intertwined. `LATEX` and `bibTEX` create parameters or change the value of parameters, and write them in an auxiliary file. The other processors may need those values to produce the correct output. The `LATEX` processor may even need the parameters in a second run. Therefore, consider the creation of the (PDF) document finished when none of the processors causes the auxiliary file to change. This is performed by a shell script `w2pdf`.

```

< make targets 59a > ≡
    nlpp.pdf : nlpp.w $(W2PDF) $(PDF_FIG_NAMES) $(PDFT_NAMES)
              chmod 775 $(W2PDF)
              $(W2PDF) $*

```

◇

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.

Fragment referenced in 54b.

Uses: pdf 58b, PDFT_NAMES 57b, PDF_FIG_NAMES 57b.

The following is an ugly fix of an unsolved problem. Currently I develop this thing, while it resides on a remote computer that is connected via the `sshfs` filesystem. On my home computer I cannot run executables on this system, but on my work-computer I can. Therefore, place the following script on a local directory.

```

< directories to create 59b > ≡
    ../nuweb/bin ◇

```

Fragment defined by 9abcd, 59b.

Fragment referenced in 64b.

Uses: nuweb 60d.

```

< parameters in Makefile 59c > ≡
    W2PDF=../nuweb/bin/w2pdf
    ◇

```

Fragment defined by 54a, 55a, 57ab, 59c, 61b, 64a.

Fragment referenced in 54b.

Uses: nuweb 60d.

```

< expliciete make regels 59d > ≡
    $(W2PDF) : nlpp.w $(NUWEB)
              $(NUWEB) nlpp.w
    ◇

```

Fragment defined by 55bd, 56ab, 58a, 59d, 61c, 62b.

Fragment referenced in 54b.

```

"../nuweb/bin/w2pdf" 59e≡
    #!/bin/bash
    # w2pdf -- compile a nuweb file
    # usage: w2pdf [filename]
    # 20170817 at 1444h: Generated by nuweb from a_nlpp.w
    NUWEB=../env/bin/nuweb
    LATEXCOMPILER=pdflatex
    < filenames in nuweb compile script 60b >
    < compile nuweb 60a >

```

◇

Uses: nuweb 60d.

The script retains a copy of the latest version of the auxiliary file. Then it runs the four processors nuweb, L^AT_EX, MakeIndex and bibT_EX, until they do not change the auxiliary file or the index.

```

⟨ compile nuweb 60a ⟩ ≡
    NUWEB=/home/paul/nlpp/nlpp/env/bin/nuweb
    ⟨ run the processors until the aux file remains unchanged 61a ⟩
    ⟨ remove the copy of the aux file 60c ⟩
    ◇

```

Fragment referenced in 59e.

Uses: nuweb 60d.

The user provides the name of the nuweb file as argument. Strip the extension (e.g. .w) from the filename and create the names of the L^AT_EX file (ends with .tex), the auxiliary file (ends with .aux) and the copy of the auxiliary file (add old. as a prefix to the auxiliary filename).

```

⟨ filenames in nuweb compile script 60b ⟩ ≡
    nufil=$1
    trunk=${1%.*}
    texfil=${trunk}.tex
    auxfil=${trunk}.aux
    oldaux=old.${trunk}.aux
    indexfil=${trunk}.idx
    oldindexfil=old.${trunk}.idx
    ◇

```

Fragment referenced in 59e.

Defines: auxfil 61a, 63ab, indexfil 61a, 63a, nufil 60d, 63ac, oldaux 60c, 61a, 63ab, oldindexfil 61a, 63a, texfil 60d, 63ac, trunk 60d, 63acd.

Remove the old copy if it is no longer needed.

```

⟨ remove the copy of the aux file 60c ⟩ ≡
    rm $oldaux
    ◇

```

Fragment referenced in 60a, 62e.

Uses: oldaux 60b, 63a.

Run the three processors. Do not use the option -o (to suppress generation of program sources) for nuweb, because w2pdf must be kept up to date as well.

```

⟨ run the three processors 60d ⟩ ≡
    $NUWEB $nufil
    $LATEXCOMPILER $texfil
    makeindex $trunk
    bibtex $trunk
    ◇

```

Fragment referenced in 61a.

Defines: bibtex 63cd, makeindex 63cd, nuweb 10e, 54a, 55bcd, 59bce, 60a, 61b, 62d.

Uses: nufil 60b, 63a, texfil 60b, 63a, trunk 60b, 63a.

Repeat to copy the auxiliary file and the index file and run the processors until the auxiliary file and the index file are equal to their copies. However, since I have not yet been able to test the aux file and the idx in the same test statement, currently only the aux file is tested.

It turns out, that sometimes a strange loop occurs in which the aux file will keep to change. Therefore, with a counter we prevent the loop to occur more than 10 times.

```

⟨run the processors until the aux file remains unchanged 61a⟩ ≡
LOOPCOUNTER=0
while
  ! cmp -s $auxfil $oldaux
do
  if [ -e $auxfil ]
  then
    cp $auxfil $oldaux
  fi
  if [ -e $indexfil ]
  then
    cp $indexfil $oldindexfil
  fi
  ⟨run the three processors 60d⟩
  if [ $LOOPCOUNTER -ge 10 ]
  then
    cp $auxfil $oldaux
  fi;
done
◇

```

Fragment referenced in 60a.

Uses: auxfil 60b, 63a, indexfil 60b, oldaux 60b, 63a, oldindexfil 60b.

A.6.4 Create HTML files

HTML is easier to read on-line than a PDF document that was made for printing. We use `tex4ht` to generate HTML code. An advantage of this system is, that we can include figures in the same way as we do for `pdflatex`.

To create a HTML doc, we do the following:

1. Create a directory `../nuweb/html` for the HTML document.
2. Put the nuweb source in it, together with style-files that are needed (see variable `HTMLSOURCE`).
3. Put the script `w2html` in it and make it executable.
4. Execute the script `w2html`.

Make a list of the entities that we mentioned above:

```

⟨parameters in Makefile 61b⟩ ≡
htmlmdir=../nuweb/html
htmlsource=nlpp.w nlpp.bib html.sty artikel3.4ht w2html
htmlmaterial=$(foreach fil, $(htmlsource), $(htmlmdir)/$(fil))
htmltarget=$(htmlmdir)/nlpp.html
◇

```

Fragment defined by 54a, 55a, 57ab, 59c, 61b, 64a.

Fragment referenced in 54b.

Uses: nuweb 60d.

Make the directory:

```

⟨expliciete make regels 61c⟩ ≡
$(htmlmdir) :
    mkdir -p $(htmlmdir)
◇

```

Fragment defined by 55bd, 56ab, 58a, 59d, 61c, 62b.

Fragment referenced in 54b.

The rule to copy files in it:

```
< implicate make regels 62a > ≡
    $(htmldir)/% : % $(htmldir)
    cp $< $(htmldir)/
```

◇

Fragment defined by 57c, 62a.

Fragment referenced in 54b.

Do the work:

```
< expliciete make regels 62b > ≡
    $(htmltarget) : $(htmlmaterial) $(htmldir)
    cd $(htmldir) && chmod 775 w2html
    cd $(htmldir) && ./w2html nlpp.w
```

◇

Fragment defined by 55bd, 56ab, 58a, 59d, 61c, 62b.

Fragment referenced in 54b.

Invoke:

```
< make targets 62c > ≡
    htm : $(htmldir) $(htmltarget)
```

◇

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.

Fragment referenced in 54b.

Create a script that performs the translation.

```
"w2html" 62d≡
#!/bin/bash
# w2html -- make a html file from a nuweb file
# usage: w2html [filename]
# [filename]: Name of the nuweb source file.
# 20170817 at 1444h: Generated by nuweb from a_nlpp.w
echo "translate " $1 >w2html.log
NUWEB=/home/paul/nlpp/nlpp/env/bin/nuweb
< filenames in w2html 63a >

< perform the task of w2html 62e >
```

◇

Uses: nuweb 60d.

The script is very much like the w2pdf script, but at this moment I have still difficulties to compile the source smoothly into HTML and that is why I make a separate file and do not recycle parts from the other file. However, the file works similar.

```
< perform the task of w2html 62e > ≡
    < run the html processors until the aux file remains unchanged 63b >
    < remove the copy of the aux file 60c >
```

◇

Fragment referenced in 62d.

The user provides the name of the nuweb file as argument. Strip the extension (e.g. `.w`) from the filename and create the names of the L^AT_EX file (ends with `.tex`), the auxiliary file (ends with `.aux`) and the copy of the auxiliary file (add `old.` as a prefix to the auxiliary filename).

```
<filenames in w2html 63a> ≡
  nufil=$1
  trunk=${1%.*}
  texfil=${trunk}.tex
  auxfil=${trunk}.aux
  oldaux=old.${trunk}.aux
  indexfil=${trunk}.idx
  oldindexfil=old.${trunk}.idx
  ◇
```

Fragment referenced in 62d.

Defines: `auxfil` 60b, 61a, 63b, `nufil` 60bd, 63c, `oldaux` 60bc, 61a, 63b, `texfil` 60bd, 63c, `trunk` 60bd, 63cd.

Uses: `indexfil` 60b, `oldindexfil` 60b.

```
<run the html processors until the aux file remains unchanged 63b> ≡
  while
    ! cmp -s $auxfil $oldaux
  do
    if [ -e $auxfil ]
    then
      cp $auxfil $oldaux
    fi
    <run the html processors 63c>
  done
  <run tex4ht 63d>
  ◇
```

Fragment referenced in 62e.

Uses: `auxfil` 60b, 63a, `oldaux` 60b, 63a.

To work for HTML, nuweb *must* be run with the `-n` option, because there are no page numbers.

```
<run the html processors 63c> ≡
  $NUWEB -o -n $nufil
  latex $texfil
  makeindex $trunk
  bibtex $trunk
  htlatex $trunk
  ◇
```

Fragment referenced in 63b.

Uses: `bibtex` 60d, `makeindex` 60d, `nufil` 60b, 63a, `texfil` 60b, 63a, `trunk` 60b, 63a.

When the compilation has been satisfied, run `makeindex` in a special way, run `bibtex` again (I don't know why this is necessary) and then run `htlatex` another time.

```
<run tex4ht 63d> ≡
  tex '\def\filename{{nlpp}{idx}{4dx}{ind}} \input idxmake.4ht'
  makeindex -o $trunk.ind $trunk.4dx
  bibtex $trunk
  htlatex $trunk
  ◇
```

Fragment referenced in 63b.

Uses: `bibtex` 60d, `makeindex` 60d, `trunk` 60b, 63a.

A.7 Perform the installation

Run nuweb, but suppress the creation of the L^AT_EX documentation. Nuweb creates only sources that do not yet exist or that have been modified. Therefore make does not have to check this. However, “make” has to create the directories for the sources if they do not yet exist. So, let’s create the directories first.

```
⟨ parameters in Makefile 64a ⟩ ≡
MKDIR = mkdir -p
```

◇

Fragment defined by 54a, 55a, 57ab, 59c, 61b, 64a.

Fragment referenced in 54b.

Defines: MKDIR 64b.

```
⟨ make targets 64b ⟩ ≡
DIRS = ⟨ directories to create 9a, ... ⟩
```

```
$(DIRS) :
$(MKDIR) $@
```

◇

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.

Fragment referenced in 54b.

Defines: DIRS 64d.

Uses: MKDIR 64a.

```
⟨ make scripts executable 64c ⟩ ≡
chmod -R 775 ../bin/*
chmod -R 775 ../env/bin/*
```

◇

Fragment defined by 7o, 8c, 17a, 24g, 36a, 46d, 64c.

Fragment referenced in 64d.

The target “sources” unpacks the nuweb file and creates the program scripts, i.e. the scripts that will apply modules on a NAF file and the script `install_modules` that installs the modules themselves and that creates the software environment the the modules need.

```
⟨ make targets 64d ⟩ ≡
sources : nlpp.w $(DIRS) $(NUWEB)
          $(NUWEB) nlpp.w
          ⟨ make scripts executable 7o, ... ⟩
```

◇

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.

Fragment referenced in 54b.

Uses: DIRS 64b.

The “install” target performs the complete installation.


```

< make targets 65a > ≡
    install : sources
             ../env/bin/make_infrastructure
             ../env/bin/install_modules

```

◇

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.

Fragment referenced in 54b.

Defines: **install** 8d, 13a, 14b, 15b, 16a, 18a, 19a, 20b, 29de, 30a, 31c, 34c, 36b, 54e, 65b.

A.8 Test whether it works

The targets `testnl` and `testen` perform the test-script to test the dutch resp. english pipeline.

```

< make targets 65b > ≡

    testnl : install test.nl.in.naf
            rm -rf ../test
            mkdir ../test
            cd ../test && ../bin/test nl

    testen : install test.en.in.naf
            rm -rf ../test
            mkdir ../test
            cd ../test && ../bin/test en

```

◇

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.

Fragment referenced in 54b.

Defines: **testen** Never used, **testnl** Never used.

Uses: **install** 65a.

A.9 Restore paths after transplantation

When an existing installation has been transplanted to another location, many path indications have to be adapted to the new situation. The scripts that are generated by nuweb can be repaired by re-running nuweb. After that, configuration files of some modules must be modified.

```

< make targets 65c > ≡
    transplant :
                touch a_nlpp.w
                $(MAKE) sources
                ../env/bin/transplant

```

◇

Fragment defined by 54d, 58b, 59a, 62c, 64bd, 65abc.

Fragment referenced in 54b.

B References

B.1 Literature

References

- [1] Rodrigo Agerri, Itziar Aldabe, Zuhaitz Beloki, Egoitz Laparra1, Maddalen Lopez de Lacalle1, German Rigau, Aitor Soroa, Antske Fokkens, Ruben Izquierdo, Marieke van Erp, Piek Vossen, Christian Girardi, and Anne-Lyse Minard. Event detection, version 2, deliverable d4.2.2. Technical report, University of the Basque Country, IXA NLP group, feb 2015. <http://www.newsreader-project.eu/files/2012/12/NWR-D4-2-2.pdf>.
- [2] Donald E. Knuth. Literate programming. Technical report STAN-CS-83-981, Stanford University, Department of Computer Science, 1983.

C Indexes

C.1 Filenames

"../bin/check_start_spotlight" Defined by 23a, 24b.
 "../bin/constpars" Defined by 46a.
 "../bin/coref-graph" Defined by 45e.
 "../bin/derel" Defined by 40b.
 "../bin/evcoref" Defined by 45b.
 "../bin/ewsd" Defined by 42g.
 "../bin/factuality" Defined by 44b.
 "../bin/framesrl" Defined by 41f.
 "../bin/heideltime" Defined by 40e.
 "../bin/m4_ukbcript" Defined by 42d.
 "../bin/mor" Defined by 38g.
 "../bin/ned" Defined by 39h.
 "../bin/nedrer" Defined by 41i.
 "../bin/nerc" Defined by 39b.
 "../bin/nlpp" Defined by 51b.
 "../bin/nomevent" Defined by 41d.
 "../bin/onto" Defined by 41b.
 "../bin/opinimin" Defined by 44e.
 "../bin/pos" Defined by 38j.
 "../bin/srl" Defined by 43a.
 "../bin/srl-dutch-nominals" Defined by 43e.
 "../bin/test" Defined by 51a.
 "../bin/tok" Defined by 38a.
 "../bin/topic" Defined by 38d.
 "../bin/wikify" Defined by 42a.
 "../bin/wsd" Defined by 39e.
 "../env/bin/chasbang.awk" Defined by 17b.
 "../env/bin/clean_infrastructure" Defined by 8a.
 "../env/bin/install_modules" Defined by 35d.
 "../env/bin/langdetect.py" Defined by 46c.
 "../env/bin/make_infrastructure" Defined by 7a.
 "../env/bin/tran" Defined by 16f.
 "../nuweb/bin/w2pdf" Defined by 59e.
 "../progen" Defined by 10a.
 "Makefile" Defined by 54b.
 "w2html" Defined by 62d.

C.2 Macro's

⟨all targets 54e⟩ Referenced in 54c.

<annotate 49b> Referenced in 51ab.
 <annotate dutch document 48b> Referenced in 49b.
 <annotate english document 49a> Referenced in 49b.
 <apply script tran on the scripts in 17c> Referenced in 16c.
 <check listener on host, port 25c> Referenced in 24b, 27c.
 <check presence of javac in 1.8 12a> Referenced in 13a.
 <check presence of maven in 3.0.5 14a> Referenced in 14b.
 <check presence of perl in 5 18e> Referenced in 19a.
 <check presence of python3 in 3.6 15a> Referenced in 15b.
 <check whether a tarball is present in the snapshot 12b> Referenced in 13a, 14b, 15b, 19a.
 <check whether XML::LibXML is installed 19f> Referenced in 19a.
 <clean up 55c> Not referenced.
 <clean up after installation 14g> Referenced in 8a.
 <compile nuweb 60a> Referenced in 59e.
 <conditionally install Perl 20a> Referenced in 7a.
 <contents of shorthand-script 36c> Referenced in 38adj, 39beh, 40be, 41bdfi, 42adg, 43ae, 44e, 45be, 46a.
 <create python script and pip script 16b> Referenced in 16a.
 <default target 54c> Referenced in 54b.
 <directories to create 9abcd, 59b> Referenced in 64b.
 <download everything 11c, 28> Referenced in 11b.
 <download stuff 13f, 18b, 21b, 29a, 33d> Referenced in 28.
 <explicitete make regels 55bd, 56ab, 58a, 59d, 61c, 62b> Referenced in 54b.
 <extract the absolute path from one of the scripts 32d> Referenced in 32c.
 <filenames in nuweb compile script 60b> Referenced in 59e.
 <filenames in w2html 63a> Referenced in 62d.
 <find a spotlightserver or exit 25a> Referenced in 51a.
 <find the nlpp root directory 10d> Not referenced.
 <function to run a module 48a> Referenced in 51ab.
 <functions of the module-installer 36b> Referenced in 35d.
 <get a testfile and set naflang or die 50> Referenced in 51a.
 <get commandline-arguments for check_start_spotlight 23c> Referenced in 23a.
 <get location of the script 52a> Referenced in 7a, 8a, 10a, 16f, 23a, 35d, 36c, 37b, 44b.
 <get spotlight language parameters 25b> Not referenced.
 <get spotlight model ball 22b> Not referenced.
 <impliciete make regels 57c, 62a> Referenced in 54b.
 <init make_infrastructure 8e, 11b> Referenced in 7a, 8a.
 <install ActivePython 16a> Referenced in 15b.
 <install Alpino 30a> Referenced in 7a.
 <install boost 35b> Referenced in 7a.
 <install libxml2 or libxslt 29e> Referenced in 29f.
 <install Perl 20b> Referenced in 20a.
 <install perl 20c, 21a> Referenced in 19a.
 <install shared libs 29f> Referenced in 7a.
 <install svmlib 35a> Referenced in 7a.
 <install the modules 37d, 38cfi, 39adg, 40ad, 41ahk, 42cfi, 43cd, 44ad, 45adg> Referenced in 35d.
 <install the Spotlight server 22ac> Referenced in 7a.
 <install the ticcutils utility 34a> Referenced in 7a, 34d.
 <install the timbl utility 34b> Referenced in 7a, 34d.
 <install the treetagger utility 31acde, 32ab, 33bc> Referenced in 7a.
 <logmess 52c> Not referenced.
 <make scripts executable 7o, 8c, 17a, 24g, 36a, 46d, 64c> Referenced in 64d.
 <make targets 54d, 58b, 59a, 62c, 64bd, 65abc> Referenced in 54b.
 <make treetagger location-independent 32c> Referenced in 32b.
 <matchscript 32e> Referenced in 32d.
 <need to wget 11d> Referenced in 13f, 18b, 21b, 29a, 33d.
 <next part 7p> Referenced in 7a.
 <parameters in Makefile 54a, 55a, 57ab, 59c, 61b, 64a> Referenced in 54b.
 <perform the task of w2html 62e> Referenced in 62d.

<re-install modules after the transplantation 34d> Not referenced.
 <remove the copy of the aux file 60c> Referenced in 60a, 62e.
 <replace the absolute paths 33a> Referenced in 32c.
 <rewrite ActivePython shabangs 16c> Referenced in 16a.
 <run in subshell when naflang is not known 47a> Not referenced.
 <run only if language is English or Dutch 47b> Not referenced.
 <run tex4ht 63d> Referenced in 63b.
 <run the html processors 63c> Referenced in 63b.
 <run the html processors until the aux file remains unchanged 63b> Referenced in 62e.
 <run the processors until the aux file remains unchanged 61a> Referenced in 60a.
 <run the three processors 60d> Referenced in 61a.
 <set default arguments for Spotlight 24a> Referenced in 23a.
 <set environment parameters 10c, 30b, 31b, 34e, 35c> Referenced in 10a.
 <set the naflang parameter 37a> Referenced in 36c, 44b.
 <set up autoconf 29d> Referenced in 7a.
 <set up Java 13a> Referenced in 7a.
 <set up java environment 13e> Referenced in 13a.
 <set up Maven 14b> Referenced in 7a.
 <set up Perl 19a> Not referenced.
 <set up Python 15b, 18a> Referenced in 7a.
 <set variables that point to the directory-structure 10ef, 11a, 14f> Referenced in 10a.
 <start of module-script 37b> Not referenced.
 <start the Spotlight server on localhost 27ab> Referenced in 24b, 26a.
 <test presence of command 8d> Referenced in 8e.
 <test whether spotlightserver runs 26e> Referenced in 26a.
 <try to obtain a running spotlightserver 26a> Not referenced.
 <unpack ticcutils or timbl 34c> Referenced in 34ab.
 <variables of the module-installer 52b> Referenced in 35d.
 <wait until the spotlight server is up or faulty 27c> Referenced in 27b.

C.3 Variables

all: 54c.
 ALPINO_HOME: 30b.
 auxfil: 60b, 61a, 63a, 63b.
 bibtex: 60d, 63cd.
 DIRS: 64b, 64d.
 fig2dev: 57c.
 FIGFILENAMES: 57b.
 FIGFILES: 57a, 57b.
 indexfil: 60b, 61a, 63a.
 install: 8d, 13a, 14b, 15b, 16a, 18a, 19a, 20b, 29de, 30a, 31c, 34c, 36b, 54e, 65a, 65b.
 makeindex: 60d, 63cd.
 MKDIR: 64a, 64b.
 moduleresult: 48a, 51ab.
 naflang: 23c, 24a, 25ab, 27a, 37a, 44b, 47ab, 49b, 50.
 nufil: 60b, 60d, 63a, 63c.
 nuweb: 10e, 54a, 55bcd, 59bce, 60a, 60d, 61b, 62d.
 oldaux: 60b, 60c, 61a, 63a, 63b.
 oldindexfil: 60b, 61a, 63a.
 opinion_models_ball_path: 44d.
 PATH: 11a, 13e, 14bf, 20c.
 pdf: 55a, 58b, 59a.
 PDFT_NAMES: 57b, 59a.
 PDF_FIG_NAMES: 57b, 59a.
 PHONY: 54c, 58a.
 piperoot: 10ab, 10d, 10e, 13e, 16ac, 20c, 29de, 34c, 35a, 36b, 51a.
 print: 17bc, 20a, 25a, 32e, 46c, 56a, 58b.

PST_NAMES: [57b](#).
PS_FIG_NAMES: [57b](#).
runmodule: [48a](#), [48b](#), [49a](#).
SUFFIXES: [55a](#).
SVMLIB_HOME: [34e](#), [35a](#).
testen: [65b](#).
testnl: [65b](#).
texfil: [60b](#), [60d](#), [63a](#), [63c](#).
TREETAGGER_HOME: [31a](#), [31b](#), [32d](#), [33a](#).
trunk: [60b](#), [60d](#), [63a](#), [63cd](#).
view: [58b](#).