
ANNIS User Guide

User Guide

Amir Zeldes <annis-admin@ling.uni-potsdam.de>

Version 2.2.1

18 April 2012

Table of Contents

Introduction	1
New Features in Version 2.2.1	1
Installing ANNIS2	2
Installing a Local Version (ANNIS Kickstarter)	2
Building and Installing an ANNIS Server	2
Running Queries in ANNIS2	4
The ANNIS2 Interface	4

Introduction

ANNIS is an open source, browser-based search and visualization architecture for multi-layer corpora. It can be used to search for complex graph structures of annotated nodes and edges forming a variety of linguistic structures, such as constituent or dependency syntax trees, coreference and parallel alignment edges, span annotations and associated multi-modal data (audio/video). This guide provides an overview of the current ANNIS system, first steps for installing either a local instance or an ANNIS server with a demo corpus, as well as tutorials for converting data for ANNIS and running queries with AQL (ANNIS Query Language).

New Features in Version 2.2.1

Features:

- Subcorpus and document path for each hit can be shown in search result list
- Metadata coming from documents, subcorpora and top level corpora is sorted separately in the metadata display
- New grid-tree visualizer for a span visualization of hierarchical structures (e.g. topological fields in German; see Section 5.1)
- More intuitive search button position in GUI and other small GUI improvements
- Updated tutorial
- Faster import when large corpora are already in the system

Bugfixes:

- Support for JDK 7
- Fixed import bug which could disrupt queries of the sort `>secedge m,n` (see <https://bugs.launchpad.net/bugs/870108>)
- Fixed bugs in some forms of query disjunction with `|`
- Fixed resizing of visualizers in Chrome browsers
- Fixed escaping of characters in WEKA export

(For change logs of previous version see their respective distributions or user guides)

Installing ANNIS2

Installing a Local Version (ANNIS Kickstarter)

Local users who do not wish to make their corpora available online can install ANNIS Kickstarter under most versions of Linux, Windows and Mac OS. To install Kickstarter follow these steps:

1. Download and install PostgreSQL 8.4 for your operating system from <http://www.postgresql.org/download/> and **make a note of the administrator password** you set during the installation. After installation, Postgres may automatically launch the Postgres Stack Builder to download additional components – you can safely skip this step and cancel the Stack Builder if you wish. You may need to restart your OS if the Postgres installer tells you to.
2. Download and unzip `Annis-Kickstarter-2.2.1-distribution.zip` [<http://launchpad.net/annis/2.2/2.2.1/+download/Annis-Kickstarter-2.2.1-distribution.zip>] from the ANNIS website.
3. Start `AnnisKickstarter.bat` if you're using Windows or run the bash script `AnnisKickstarter.sh` otherwise (this may take a few seconds the first time you run Kickstarter). At this point your Firewall may try to block Kickstarter and offer you to unblock it – do so and Kickstarter should start up.



Note

For most users it is a good idea to give Java more memory (if this is not already the default). You can do this by editing the script `AnnisKickstarter` and typing the following after the call to start java (before `-splash:splashscreen.gif`):

```
-Xss1024k -Xmx1024m
```

(To accelerate searches it is also possible to give the Postgres database more memory, see the link in the next section below).

4. Once the program has started, if this is the first time you run Kickstarter, press “Init Database” and supply your PostGres administrator password from step 1.
5. Download and unzip the pcc2 demo corpus [http://korpling.german.hu-berlin.de/~annis/downloads/sample_corpora/pcc2_relAnnis.zip] from the ANNIS website.
6. Press “Import Corpus” and navigate to the directory containing the directory `pcc2_v2_relAnnis/`. Select this directory (but do not go into it) and press OK.
7. Once import is complete, press “Launch Annis frontend” and login with the username and password “test” to test the corpus (try selecting the pcc2 corpus, typing `pos="NN"` in the AnnisQL box and clicking “Show Result”. See the section “Running Queries in ANNIS2” in this guide for some more example queries, or press the Tutorial button at the top left of the interface).

Building and Installing an ANNIS Server

The ANNIS server version can be installed on UNIX based server, or else under Windows using Cygwin [<http://www.cygwin.com/>], the freely available UNIX emulator. To install the ANNIS server:

1. Install a PostgreSQL server for your operating system from <http://www.postgresql.org/download/>
2. Install a web server such as Tomcat [<http://tomcat.apache.org/>] or Jetty [<http://www.mortbay.org/jetty/>]
3. Make sure you have JDK 6 [<http://java.sun.com/javase/downloads/index.jsp>] and Maven 2 [<http://maven.apache.org/>] (or install them if you don't)

4. If you're using Cygwin and Windows you will also need to install the "patch" program via the Cygwin package manager

5. Download and unzip Annis-2.2.1.zip [<http://launchpad.net/annis/2.2/2.2.1/+download/Annis-2.2.1.zip>], then run the following commands (replacing the appropriate directories):

```
cd <unzipped source>/Annis-Service
mvn -DskipTests=true install
mvn -DskipTests=true assembly:assembly tar xzvf
target/annis-service-<version>-distribution.tar.gz -C <installation directory>
```

6. Next initialize your ANNIS database (only the first time you use the system):

7. Set the environment variables (each time when starting up)

```
export ANNIS_HOME=<installation directory>
export PATH=$PATH:$ANNIS_HOME/bin
```

8. Now you can import some corpora:

```
annis-admin.sh import path/to/corpus1 path/to/corpus2 ...
```



Important

The above import-command calls other PostgreSQL database commands. If you abort the import script with Ctrl+C, these SQL processes will not be automatically terminated; instead they might keep hanging and prevent access to the database. The same might happen if you close your shell before the import script terminates, so you will want to prefix it with the "nohup"-command.

9. Now you can start the ANNIS service:

```
annis-service.sh start
```

10. To get the Annis front-end running, first compile it:

```
cd <unzipped source>
mvn -DskipTests=true install
```

If no error occurs the war-file will be available under `<unzipped source>/Annis-web/target/Annis-web.war`.

11. And configure your web server as described here: <https://korpling.german.hu-berlin.de/p/projects/annis/wiki/Tomcat> [<https://korpling.german.hu-berlin.de/p/projects/annis/wiki/Tomcat>]

The latest instructions for compiling and installing the ANNIS Server can also be found at: <https://korpling.german.hu-berlin.de/p/projects/annis/wiki/Documentation> [<https://korpling.german.hu-berlin.de/p/projects/annis/wiki/Documentation>]

We also **strongly recommend** reconfiguring the Postgres server's default settings as described here: <https://korpling.german.hu-berlin.de/p/projects/annis/wiki/PostgreSQL>

Running Queries in ANNIS2

The ANNIS2 Interface

Figure 1. ANNIS2 interface

The screenshot displays the ANNIS2 interface with two main panels. The left panel, titled 'Search Form', contains an 'AnnisQL' input field with a complex query, a 'Result:' field showing 'Valid Query', a table of 'More Corpora' with checkboxes, and search parameters like 'Context Left: 0', 'Context Right: 0', and 'Results per page: 10'. The right panel, titled 'Search Result', shows the results for the query 'tok & tok & #1 -> dep[func="OA"] #2 & cat="S" & #3 _#1 & node & #3 > secedge #4 (0, 0)'. It includes a video player, a list of tokens with their grammatical functions and categories, dependency and constituent trees, and a list of annotations with their levels and types.

Search Form

AnnisQL: tok & tok & #1 -> dep[func="OA"] #2 & cat="S" & #3 _#1 & node & #3 > secedge #4 | correction="correcting" | cat="S"

Result: Valid Query

More Corpora

Name	Texts	Tokens
FalkoEssayL2V2_0	248	131511
ONTONOTES_v1.5_small	4	6450
SMULTRON_Banana	2	3782
TueBa5_no_cyc	2187	770949
agni_I	24	184
b4.tatian2.0	2031	11295
pcc-3	3	573
pcc2	2	399
<input checked="" type="checkbox"/> tiger1_dep	1	929
<input type="checkbox"/> tiger2	1971	888578

Search Export

Context Left: 0

Context Right: 0

Results per page: 10

Show Result

Search Result

Page 1 of 5

Token Annotations Show Citation URL

Displaying Results 1 - 10 of 43

während 78 Prozent sich für Bush und vier Prozent für Clinton aussprechen

KOUS CARD NN PRF APPR NE KON CARD NN APPR NE VVFIN

-- -- *.Neut 3.Acc.PI -- -- Acc.Sq.* -- -- *.Neut -- -- Acc.Sq.* 3.PI.Past.Ind

dependencies

constituents

Die Vase auf dem Tisch ist größer als die Vase

animacy (grid)

Select Displayed Annotation Levels

mmaxref_type	inanim	inanim	inanim
mmaxref_type			
tok	Die	Vase	auf dem Tisch ist größer als die Vase

coreference (discourse)

Die Vase auf dem Tisch ist größer als die Vase auf der Fensterbank . Ich finde sie sieht nicht so gut aus , weil der Tisch zu klein ist .

The ANNIS2 interface (see Figure 1, “ANNIS2 interface”) is comprised of several windows, the most important of which are the search form (in the red box above left) and the results window (in the blue box above).

The Search Form

The Search Form on the left of the interface window is available immediately after login. In the middle, the list of currently available corpora is shown. Using the checkboxes on the left of each corpus, it is possible to select which corpora should be searched in (hold down 'shift' to select multiple corpora simultaneously). If you cannot see a corpus that should be available to you, or else if the corpora list is too cluttered, you may click on "more corpora" to open the corpora window. You may then drag and drop the des