Short Introduction to the ANNIS Query Language

Potsdam, July 3rd 2007; Draft version.

1 What you can search for

The *ANNIS Query Language (AQL)* enables you to search for the following information:

- *Documents*
- *Text*
- *Annotations*

2 Document Search

- To search for a document, you can use the *document name* .
- Underspecified queries can be formulated by combining the *wildcards* ***** and **?** with the *document name*.

Pattern:

doc=*DocumentName*

Examples:

doc=urml.maz-10374.anno

doc=urml*

3 Text Search

There are two ways to search for the occurrences of some text:

- Using a *text string* enclosed by *double quotation marks*
- Using *Regular Expressions* enclosed by two slashes

Pattern:

"*TextString*" (it's possible to use *wildcards* (* ?) as placeholders within the *text string*)

/*RegExp*/

Examples:

"Jugendliche"

"Musikcaf?"

/Fußball.+/

4 Annotation Search

To search for *annotations,* you can use the following information:

- *(name of the) Annotation Tag set (ATS)*
- *Annotation Attribute*
- *Annotation Value*
- *Text*

The query expression puts this elements in a fixed order.

But it's possible to replace some of the elements by *wildcards* to underspecify your query.

(At least one of the elements has to be a *non-wildcard* term!)

Pattern:

ATS::Attribute=Value="Text"

***partial expressions (unique)*:**	***full expression*:**
ATS::Attribute=Value	ATS::Attribute=Value=*
ATS::Attribute="Text"	ATS::Attribute=*="Text"
Attribute=Value="Text"	*::Attribute=Value="Text"
Attribute=Value	*::Attribute=Value=*
ATS::"Text"	ATS::*=*="Text"

Examples:

pcc11-coref::grammatical_role=SBJ

pcc11-tiger::cat="*Schröder*"

pos=NN="K*"

Foc_contr=cf-conf

pcc11-rst::"kein*"

5 Complex search terms

Elementary search terms of each type (*Document, Text, Annotation*) can be combined to form complex query expressions by the following *operators*:

- **&** (means *AND THERE IS*)
- **|** (means *OR*)
- **^** (means *exclusive OR*)

The *precedence* (of order of interpretation) corresponds to the order of presentation(decreasingly).

You can use *brackets* to force another order of interpretation or to increase the readability.

Pattern:

Term1 **&** *Term2*

Term1 | *Term2*

Term1 **^** *Term2*

(Term1) (Term2) * same as: *Term1* **&** *Term2*

***Examples**:*

inf-stat=new | inf-stat=acc*

6 Extended Match Conditions

In particular with the *Annotation Search*, there are some extended *relational match conditions*.

To query for a special relation between two terms *a* and *b*, you first have to list all the terms (of the expression):

a & b & c (first part)

The *second part* specifies the desired relation, identifying the related terms by there position:

#1 REL-OP #2 (second part)

To obtain the complete expression, part *one* and *two* have to be connected via the operator '&':

a & b & c & #1 REL-OP #2

Pattern:

term1 & term2 & #1 REL-OP #2

List of operators:

Symbol	***Description***	**Example**
a > *b*	*a* directly dominates *b*	**s** & np & #1 > #2
a **>*** *b*	*a* dominates *b*	**s** & pos=ADJA & #1 >* #2
a **$** *b*	*a* is a sister node of *b*	pos=art & pos=ADJA & #1 $ #2
a **.** *b*	*a* directly precedes *b*	art & adja & #1 . #2
a **.*** *b*	*a* precedes *b*	art & adja & #1 .* #2
a **_=_** *b*	*a* covers the same sequence as *b*	np & n & #1 _=_ #2
a **@** *b*	*a* and *b* are types/values of the same ATS	n & np & #1 @ #2
a:root	is the root node tree	s & #1:root
a:arity=*n*	*a* has *n* daughter nodes	s & #1:arity=3

Christian Lemke, Michael Götze; July 3rd 2007