

Hilfe zur Suche in ANNIS 1.1

– DRAFT –

SFB 632, D1
annis@ling.uni-potsdam.de

27. August 2004

1 Die Suche

Thema dieses Abschnittes ist

- das Formulieren von Suchanfragen in ANNIS
- die Darstellung der Suchergebnisse
- sowie das Speichern und Laden ausgewählter Suchergebnisse.

Die Beispiel-Suchanfragen in diesem Abschnitt sind so gewählt, dass der Benutzer sie selbst eingeben und aktiv nachvollziehen kann (und das auch tun sollte!). Die Ausdrücke, die der Benutzer einzugeben hat, sind jeweils fett gedruckt.

Auf der linken Seite des ANNIS Bildschirms (in der Menüleiste) befindet sich das Suchfenster. In diesem Fenster werden die Anfragen formuliert; durch Drücken des “Los”-Buttons (oberhalb der Suchmaske) wird der Suchprozess gestartet. Die Suchergebnisse werden im Hauptfenster (rechts) angezeigt.

Hinweis: Sie können mit den Oben-/Unten-Pfeilen (links vom “Los”-Button) in Ihren alten Anfragen blättern und eine Anfrage erneut abschicken.

1.1 Einfache Suche

Die allgemeine Form einer einfachen Suchanfrage lautet: “**Text**” oder **Attribut=Wert**. Diese Form wird im Folgenden genauer erklärt.

1.1.1 Suche nach Text

Wir beginnen mit der Suche nach einzelnen Wörtern, Wortfragmenten oder ganzen Textpassagen. Dazu wird der gesuchte Ausdruck, umgeben von doppelten Anführungszeichen “...”, in das Suchfenster eingegeben und anschließend der Suchprozess durch Drücken des “Los”-Buttons gestartet.

Beispiel 1 Wir suchen nach Texten, die den Namen “Johannes” enthalten. Geben Sie das gesuchte Wort in die Suchmaske ein: **“Johannes”**, und drücken Sie auf den “Los”-Button. Nach kurzer Zeit erscheinen im Hauptfenster die Suchergebnisse, die sogenannten ‘Treffer’.



Abbildung 1: Eine Suchanfrage

In der ersten Zeile wird die Anzahl der Treffer angegeben. In unserem Beispiel sollten Sie 1 Treffer erhalten. Das Dokument, das den Namen “Johannes” enthält, heißt “maz6046_anno”. Direkt unter dem Dokumentnamen wird der Textanfang dieses Dokuments gezeigt: “Veronas Tränen ...”. Darunter kommt dann die Auflistung der Treffer, in unserem Fall nur einer: das Wort “Johannes”, violett hinterlegt, gefolgt von einer Positionsangabe (die Buchstaben des Wortes “Johannes” nehmen die Positionen 261-268 im Dokument ein).

Die Treffer fungieren als Links, d.h. durch Klicken auf den Link können Sie sich das gesuchte Wort im Kontext anzeigen lassen. Klicken Sie nun auf “Johannes”, dann erhalten Sie den Kontext “. . . Und Johannes B. Kerner . . .”. Im Text ist das Wort “Johannes” fett unterstrichen: die Unterstreichung markiert den Ausdruck, nach dem gesucht wurde.

Sie können nun entweder direkt eine neue Suchanfrage in das Suchfenster eingeben oder aber Sie kehren mit dem “zurück”-Knopf Ihres Browsers zur Trefferauflistung zurück.



Abbildung 2: “Johannes” im Kontext

Hinweis: Groß- und Kleinschreibung wird momentan noch nicht unterschieden. Sie erhalten also dieselben Treffer, wenn Sie **“johannes”** eingeben.

Beispiel 2 Wir suchen nun nach dem Wort “erst”. Geben Sie dazu den entsprechenden Suchausdruck ein: **“erst”**. Sie sollten eine Reihe von Treffern erhalten, z.T. sogar mehrere innerhalb eines Dokuments. Z.B. finden sich im Dokument “maz-feigenblatt-11299” (bzw.: “maz11299_anno”) 2 Treffer. Wählen Sie den ersten davon aus (“erst 91.94”) und lassen Sie sich den Kontext dazu anzeigen (durch Klicken auf “erst”). Der dazugehörige Kontext ist “bei der ersten Zossener Runde” — ein wohl unerwarteter Treffer. Es werden also nicht nur Instanzen des Wortes “erst” gefunden, sondern auch Ausdrücke, in denen “erst” enthalten ist, z.B. “ersten” oder “Verständnis” oder “widersteht”.

Um die anderen Treffer anzuschauen, haben Sie zwei Möglichkeiten: Entweder Sie machen (wie oben erwähnt) Gebrauch vom “zurück”-Knopf Ihres Browsers und kehren damit zur Trefferauflistung zurück. Oder Sie benutzen spezielle Pfeile, die zum Blättern zwischen den einzelnen Treffern gedacht sind: Oben rechts im Hauptfenster (rechts neben den normalen scroll-Pfeilen) erscheinen, sobald es mehr als einen Treffer gibt, Pfeile nach rechts

und links. Klicken Sie auf den rechten Pfeil, um zum nächsten Treffer zu gelangen (“der Stadt dafür Verständnis haben”), oder auf den linken Pfeil, um zum vorhergehenden Treffer zu kommen. (Der Pfeil nach unten dient zum Speichern ausgewählter Treffer. Mehr dazu s. unten.)

Wie kann man nun die Anfrage so modifizieren, dass tatsächlich nur das Wort “erst” gefunden wird? Das gewünschte Ergebnis können Sie erreichen, indem Sie vor und hinter dem gesuchten Begriff ein Leerzeichen einfügen: “**erst**”. Jetzt sollten Sie deutlich weniger Treffer erhalten.

Man kann sich die besondere Eigenschaft der Textsuche auch zu Nutze machen: z.B. können auf diese Weise alle Zusammensetzungen mit “Fußball” gefunden werden. Geben Sie den Suchausdruck “**Fußball**” ein. Sie sollten dann neben dem Wort “Fußball” auch “Fußball-Weltmacht” und “Fußball-Kleinmacht” finden (im Dokument “maz-furchtet-10374”).

1.1.2 Suche nach Annotationen

Die Texte in ANNIS sind mit verschiedener linguistischer Information annotiert, wie z.B. Informationsstruktur, Syntax etc. In den meisten Fällen besteht die annotierte Information aus einem Attribut mit einem Wert, wie z.B. dem Attribut “case” mit dem Wert “nominative”.

Neben reinem Text können Sie auch in diesen Annotationen suchen. Hier lautet der allgemeine Suchaufbau **Attribut=Wert**, also z.B. **case=nominative**.

Beispiel 3 Wir suchen nach Topikausdrücken. In unseren Beispieldokumenten heißt das relevante Attribut “Topic”, ein möglicher Wert dieses Attributs ist “aboutness-topic”. Daher lautet die Suchanfrage **topic=aboutness-topic**. (Beachten Sie, dass der Suchausdruck hier ohne Anführungszeichen eingegeben wird — dies unterscheidet die Suche nach Annotationen von der Suche nach Text.)

Sie sollten ca. 180 Treffer erhalten. Klicken Sie nun z.B. auf den ersten Treffer im Dokument “maz-alles-16153”: “In Berlin”. Sie gelangen damit, wie immer, zur Kontextansicht. Wenn Sie nun auf den Dreiecksschalter hinter dem Wort “Annotationen” der Annotationsebene *pcc10.is.aboutness-topic* klicken, so werden die zum Text gehörenden Annotationen mit angezeigt: “In Berlin” ist annotiert mit der Information “medium” (was die Konfidenz der Annotation angibt) und “aboutness-topic” (= dem Wert für das Attribut “topic”). Der Wert “aboutness-topic” ist dabei fett unterstrichen, da nach dieser Annotation gesucht wurde.

Fahren Sie nun mit dem Mauszeiger auf die Annotation “aboutness-topic” und warten Sie, bis der dazugehörige Informationstext erscheint, der die vollständige Annotation (inklusive des Attributs) anzeigt: “topic=aboutness-topic [4..5]”. Wir sehen also, dass der Informationstext der Suchanfrage entspricht. (“4..5” bedeutet: die Wörter “In Berlin”, die als Topik annotiert

sind und daher den Treffertext darstellen, sind die Wörter 4-5 im Dokument. Der linke Kontext, “Alles offen .”, sind die Wörter 1-3.)



Abbildung 3: Suche nach Annotationen

1.2 Kombinierte Suche

Bisher haben wir in unseren Suchanfragen immer nur genau eine Suchbedingung angegeben: ein Text(fragment) oder eine Annotation.

Will man mehrere Bedingungen innerhalb einer Suche kombinieren, so werden zuerst die einzelnen Bedingungen aufgezählt, getrennt durch “&”: **Bedingung1 & Bedingung2** Anschließend können die Konditionen mit Hilfe verschiedener Operatoren kombiniert werden. Dabei wird auf die Bedingungen der Reihe nach mit #1 (= entspricht Bedingung1), #2 etc. Bezug genommen. Die allgemeine Form der Suchanfrage lautet damit: **Bedingung1 & Bedingung2 & ... & #1 Operator #2 &**

Beispiel 4 In dem folgenden Beispiel suchen wir nach mehreren Wörtern, z.B. nach “nur noch”. Sie können entweder, wie oben gezeigt, direkt nach diesem Textstring suchen: **“nur noch”**. Ihre Anfrage sollte 2 Treffer erzielen.

Alternativ können Sie auch mit Hilfe des Präzedenz-Operators “.” fordern, dass der Ausdruck “nur” unmittelbar vor dem Ausdruck “noch” stehen soll. Die Suchanfrage lautet dann: **“nur” & “noch” & #1 . #2**. Wir suchen also nach dem Wort “nur” und dem Wort “noch”, wobei #1 “nur” entspricht und #2 “noch”. Der “.”-Operator fordert, dass das Wort “nur” dem Wort “noch” unmittelbar vorangeht.

Um die Zuordnung von #1 und #2 zu verdeutlichen, können wir die Bedingungen in der Suchanfrage auch umkehren: Geben Sie dazu ein: **“noch” & “nur” & #2 . #1**. D.h. der zweite Ausdruck soll nun vor dem ersten stehen. Jetzt sollten Sie das gleiche Ergebnis wie vorher im Hauptfenster angezeigt bekommen (2 Treffer).

Hinweis: Der “.”-Operator fordert unmittelbare Präzedenz. Sie können aber auch nach indirekter, mittelbarer Präzedenz suchen. Wenn Sie z.B. nach Sätzen suchen, die eine “je ...desto”-Konstruktion enthalten, können Sie das folgendermaßen tun: **“je” & “desto” & #1 .* #2**. “.*” erlaubt einen beliebigen Abstand zwischen den beiden Ausdrücken — Sie können ja auch nicht vorher wissen, wie groß der Abstand zwischen “je” und “desto” in den Beispielen tatsächlich ist.

Beispiel 5 Analog zur kombinierten Suche nach mehreren Wörtern können Sie auch nach mehreren Annotationen suchen. Dazu können Sie die Form der obigen Suchanfrage einfach übertragen. Beispielsweise können Sie nach einem kognitiv inaktiven Element suchen, dem unmittelbar ein Topik folgt. In unseren Beispieldokumenten heißen die entsprechenden Attribute und Werte: “cognitive_status=inactive” und “topic=aboutness-topic”. Die Anfrage würde also wie folgt lauten: **cognitive_status=inactive & topic=aboutness-topic & #1 . #2**. Sie sollten 10 Treffer erhalten. Klicken Sie z.B. auf den Treffer “von Gerhard Schröder am Samstag abend”, um zur Kontextansicht zu gelangen. Um die Annotationen zum kognitiven Status und Topik anzuschauen, müssen Sie wieder auf die entsprechenden Dreiecksschalter hinter “Annotations” klicken, in unserem Fall die von *pcc10.is.cognitive-status* und *pcc10.is.aboutness-topic*. Nun sollten Sie sehen, dass “von Gerhard Schröder” als “inactive” annotiert ist (hervorgehoben durch die fette Unterstreichung) und “am Samstagabend” als “topik”.

Beispiel 6 Alternativ können Sie auch nach mehreren Annotationen eines einzigen Ausdrucks suchen, z.B. nach einem Ausdruck, der gleichzeitig als inaktiv wie auch als Topik annotiert ist. Dafür machen wir Gebrauch vom Abdeckungsähnlichkeit-Operator “_=_”, der verlangt, dass die genannten Konditionen für genau denselben Textstring gelten müssen. Die Anfrage lautet dann: **cognitive_status=inactive & topic=aboutness-topic & #1 _=_ #2**. Hier erhalten Sie weit mehr Treffer als oben: 85.

Beispiel 7 Es ist auch möglich, die Suche nach Text mit Suche nach Annotationen zu kombinieren. Nehmen wir an, wir suchen nach dem Wort “die”, das als Subjekt (“rel=sb”) fungiert. Dafür machen wir wieder Gebrauch vom Abdeckungsgleichheit-Operator “_=_”: **“die” & rel=sb & #1 _=_ #2**. Die Anfrage verlangt, dass das Wort “die” und die Annotation als Subjekt an der gleichen Position im Korpus stehen. Für diese Anfrage sollten Sie 6 Treffer erhalten.

Hinweis für besonders Interessierte: Man kann die Anfrage auch unterspezifiziert lassen und nur eine Abdeckungsüberschneidung (statt Abdeckungsgleichheit) fordern. Dazu lässt man die 3. Kondition einfach weg: **“die” & rel=sb**. Mit dieser Anfrage erhalten Sie dann z.B. auch Treffer wie “die ewige Besserwisserei der Alten” (= 2. Treffer im Dokument “maz10205_anno”). Hier ist “die” nur (echter) Bestandteil eines Subjekts.

Beispiel 8 Für die Suche in syntaktischen Annotationen gibt es einen weiteren Operator, den Dominanz-Operator “>”. Im folgenden Beispiel suchen wir einen Satz (“cat=S”), der einen anderen Satz unmittelbar dominiert. Wieder werden zuerst die Einzelbedingungen genannt, die anschließend durch den Operator “>” miteinander verknüpft werden: **cat=S & cat=S & #1 > #2**. Sie sollten 22 Treffer erzielen.

Der Dominanz-Operator erlaubt auch eine Spezifizierung der Funktion: “>func”. Sucht man z.B. nach eingebetteten Sätzen, die als Objekt (“rel=oc”) fungieren, lautet die Anfrage: **cat=S & cat=S & #1 >oc #2**.

Hinweis für besonders Interessierte: Die Operatoren zur Suche in Bäumen sind von TIGERSearch entlehnt, s. <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/>.

Beispiel 9 Wenn Sie wie oben nach Text oder Annotationen suchen, so durchsucht ANNIS stets alle Texte in der Datenbank. Sie können die Suche aber auch auf ein oder mehrere Dokumente beschränken.

Als Beispiel nehmen wir die Anfrage aus Beispiel 7: **“die” & rel=sb & #1 _=_ #2**. (Zur Erinnerung: Sie können mit den Oben-/Unten-Pfeilen links des “Los”-Buttons in Ihren alten Anfragen blättern und eine Anfrage erneut abschicken.) Wenn Sie die Anfrage erneut starten, erhalten Sie 6 Treffer aus 2 verschiedenen Dokumenten. Jetzt können Sie die Suche weiter einschränken, indem sie den Suchraum auf ein Dokument beschränken, z.B. auf das Dokument maz10205_anno. Der Dokumentname kann als eine Art von Annotation betrachtet werden; das Attribut heißt “doc”, der Name des Dokuments bestimmt den Wert des Attributs: “doc=maz10205_anno”.

Die Suchanfrage für dieses Beispiel lautet dann: **“die” & rel=sb & #1 _=_ #2 & doc=maz10205_anno**. Jetzt werden nur noch Treffer innerhalb dieses Dokuments gefunden (3 Treffer).



Abbildung 4: Trefferliste

1.3 Speichern und Laden ausgewählter Suchergebnisse

Oft möchte man bestimmte Treffer abspeichern und zu einem späteren Zeitpunkt wieder ansehen. Das ist möglich in ANNIS: Als Beispiel nehmen wir wieder die Suchanfrage aus Beispiel 7: **“die” & rel=sb & #1 =_ #2**.

Wir erhalten im Hauptfenster die Liste der Treffer, welche violett hinterlegt sind. Klicken Sie den ersten der Links im Dokument “maz10205_anno” an, um zur Kontextansicht zu gelangen: “die 69..71”.

Wie schon erwähnt, gibt es oben rechts im Hauptfenster einen Button (ein Pfeil nach rechts), um in den verschiedenen Treffern zu blättern. Klicken Sie auf den Pfeil, um zum zweiten Treffer zu gelangen. Jetzt können Sie entweder zum dritten Treffer weitergehen (Pfeil nach rechts) oder zum ersten zurückkehren (Pfeil nach links).

Links von den Pfeilbuttons gibt es einen Button mit einem Pfeil nach unten, der zum Speichern des aktuellen Treffers dient. Klicken Sie den Pfeil an, dann wird Ihnen die ID genannt, unter der der Treffer abgespeichert wird. (Um weitere Treffer abzuspeichern, kehren Sie mit dem “zurück”-Knopf Ihres Browsers zur Detailansicht zurück und blättern Sie zu weiteren Treffern.)

Wenn Sie zu einem späteren Zeitpunkt auf die abgespeicherten Treffer zugreifen wollen, klicken Sie auf den Link “Trefferliste” (oberhalb des Suchfensters). Die bisher gespeicherten Treffer werden nun im Hauptfenster

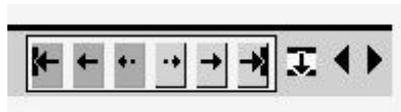


Abbildung 5: Pfeilbuttons: zum Navigieren, zum Abspeichern der Treffer und zum Blättern in den Treffern

aufgelistet. In der ersten Spalte steht die ID des Treffers, in der dritten der Name des Dokuments, das den Treffer enthält. Spalte 4 gibt den Zeitpunkt des Abspeicherns an, Spalte 5 den Suchausdruck, mit der der entsprechende Treffer gefunden wurde.

Durch Klicken auf die ID gelangen Sie zur Kontextansicht des Treffers. Ein Klick auf das rote Kreuz rechts neben dem Treffer löscht den Eintrag.