

# The COBHUNI Corpus

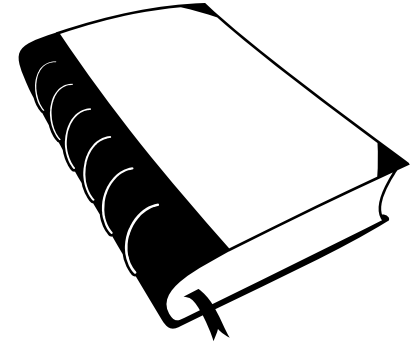


Alicia González and Tillmann Feige

# Outline

- 1 The texts
- 2 The tagset
- 3 The annotation: WebAnno
- 4 Search and visualization: Annis
- 5 Sum-up: technical workflow
- 6 Future work

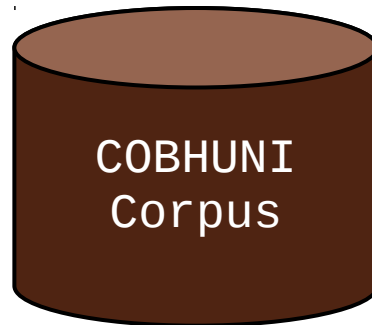
# The texts



crawl and extract

crawl and extract

scan and OCR&correct/type



# The texts

Source material	No. tokens	No. types
<i>altafsir.com</i> (Quran exegesis)	12,601,880	466,145
<i>hadith.al-islam.com</i>	11,482,139	313,846
Ocred/types texts	36,132	8,038
<b>TOTAL</b>	<b>24,120,151</b>	

# The texts

Source material	No. tokens	No. types
<i>altafsir.com</i> (Quran exegesis)	12,601,880 / <b>1,306,233</b>	466,145 / <b>125,024</b>
<i>hadith.al-islam.com</i>	11,482,139 / <b>144,122</b>	313,846 / <b>22,608</b>
Ocred/types texts	<b>36,132</b>	<b>8,038</b>
<b>TOTAL</b>	24,120,151 / <b>1,486,487</b>	

**COBHUNI**

# The tagset

## 1 MOTIVES

### 1.1 Before the unborn

- Semen and similarity / heredity
- Semen as colors
- Semen and coitus interruptus or contraceptives
- Semen and wet dream
- Sex act itself & its timing

### 1.2 The unborn

- Conception / fertilization
- General / larger debate about predestination
- Embryology: 40 days
- Embryology: Ensoulment
- Embryology: Angel visits Embryo
- Embryology: expressed in a series of numbers
- Embryology: Macrocosm - microcosm
- Embryology: Embryo and link to resurrection & afterlife
- Embryology: Link to (modern) science
- Pregnancy: duration: Definition

### 1.3 After the Unborn

- Miscarriage / abortion and legal status of slave mother
- Miscarriage / abortion and legal status of free mother
- Miscarriage / abortion and legal status of the siqt
- Abortion compared to killing a new-born
- Menstruation
- Breast-feeding
- Legal status questions concerning the child after birth

## 2 METAMOTIVES

- Philological exegesis
- Hadith criticism
- Latin script

## 3 NAMED ENTITIES → Proper name

# The tagset

## 1 MOTIVES

### 1.1 Before the unborn

- Semen and similarity / heredity
- Semen as colors
- Semen and coitus interruptus or contraceptives
- Semen and wet dream
- Sex act itself & its timing

### 1.2 The unborn

- Conception / fertilization
- General / larger debate about predestination
- Embryology: 40 days
- Embryology: Ensoulment
- Embryology: Angel visits Embryo
- Embryology: expressed in a series of numbers
- Embryology: Macrocosm - microcosm
- Embryology: Embryo and link to resurrection & afterlife
- Embryology: Link to (modern) science
- Pregnancy: duration: Definition

### 1.3 After the Unborn

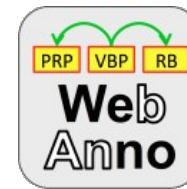
- Miscarriage / abortion and legal status of slave mother
- Miscarriage / abortion and legal status of free mother
- Miscarriage / abortion and legal status of the siqt
- Abortion compared to killing a new-born
- Menstruation
- Breast-feeding
- Legal status questions concerning the child after birth

## 2 METAMOTIVES

- Philological exegesis
- Hadith criticism
- Latin script

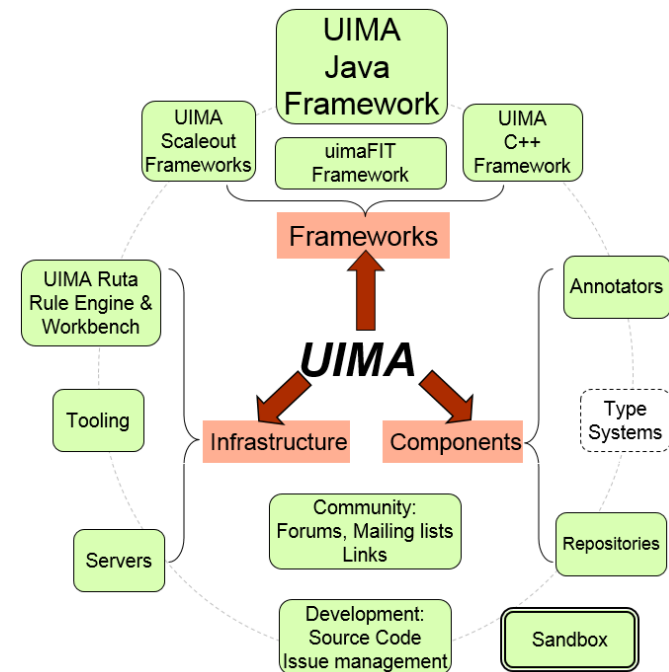
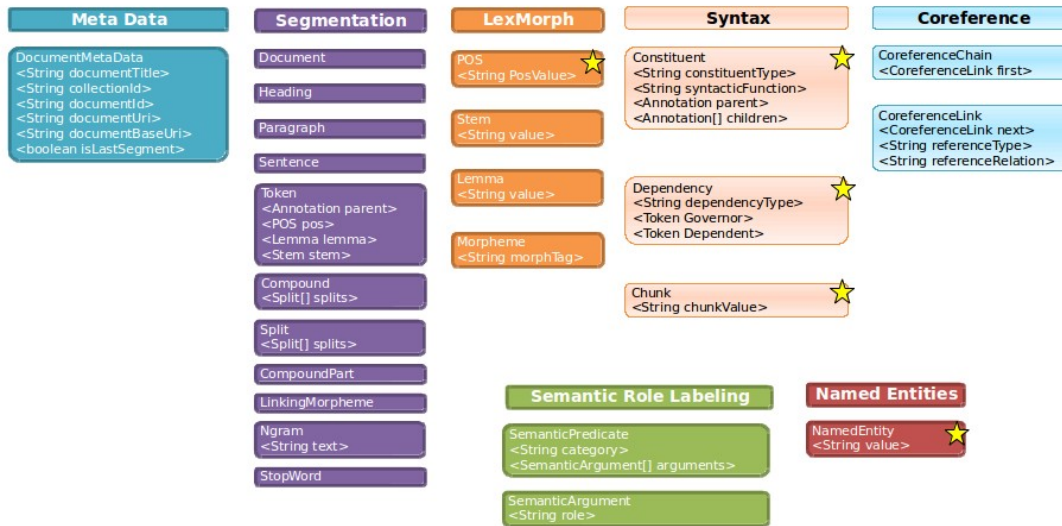
## 3 NAMED ENTITIES → Proper name

# The annotation: WebAnno



- ✓ Powerful annotation component (multi/subtoken, multilayer, overlaps)
- ✓ Import and export flexibility
- ✓ Proper visualization of Arabic script and RTL support
- ✓ Excellent support from developers team
- ✓ Robust framework (DKPro Core, Apache UIMA)

## DKPro Core Type System (Top Level)





# The annotation: WebAnno



Annotation

WebAnno | Home

Help | User: admin | Log out | Auto-logout in 29:31

Document: Open, Prev, Next, Export, Settings

Page: First, Prev, Go to, Next, Last

Script: LTR/RTL

Help: Guidelines

Workflow: Done

checkproject/Farhud\_11000\_01\_X\_1.xmi

showing 1-30 of 147 sentences [document 4 of 11]

Embryology: 40 days    Conception / fertilization    PER

1 خلق البشر عن أبي عبد الرحمن عبد الله بن مسعود - رضي الله تعالى عنه - قال: حدثنا رسول الله - صلى الله عليه وسلم - وهو الصادق المصدوق: "إن أحدكم يجمع خلقه في بطن أمه أربعين يوماً

Embryology: Ensoulment    Embryology: Angel visits Embryo    Qur'anic terminology    Qur'anic terminology    Qur'anic terminology

نطفة ، ثم يكون    علقة    مثل ذلك، ثم يكون    مضغمة    مثل ذلك، ثم    يرسل إليه الملك    فينفخ فيه الروح ، ويؤمر بأربع كلمات:    Embryology: Angel visits Embryo

يكتب رزقه، وعمله، وأجله، وشقي أو سعيد.

General / larger debate about predestination (Qadar)

2 فولذي لا إله غيره إن أحدكم ليعمل بعمل أهل الجنة حتى ما يكون بينه وبينها إلا ذراع فيسبق عليه الكتاب فيعمل بعمل أهل النار حتى ما يكون بينه وبينها إلا ذراع فيسبق عليه الكتاب فيعمل ؛

3 راوي الحديث: هو عبد الله بن مسعود، الصحابي الجليل، وسادس من أسلم، وأمين سر الرسول وصاحب سادته وتعليه وطهوره في السفر، وأبوه مسعود بن غافل ابن حبيب بن شمع بن فارس بن مخزوم بن صاهلة بن كاهل بن الحارث بن تميم بن سعد بن هذيل.

4 وكان ابن مسعود عندما أسلم، راعيا لعقبة بن معيط ولقى العنت من المشركين، وهاجر إلى الحبشة الهجرتين، وبعد هجرته إلى المدينة أذى الرسول بينه وبين الزبير بن العوام، وشهد ابن مسعود بدرًا وسائر المشاهد مع رسول الله، وفي بدر أتيح له أن يجهز على أبي جهل بعد أن أسقط جريحا.

5 كان ابن مسعود شديد السمرة، خفيف اللحم، قصير القامة، دقيق الساقين، وكان من أجود الناس ثوبًا، وأطيبهم ريحًا.

6 وقد لازم الرسول الكريم منذ إسلامه، حتى قال فيه أبو موسى الأشعري: رأيت رسول الله وما أرى إلا أن ابن مسعود من أهل بيته.

7 كان يكثر الصلاة ويقل من صيام التطوع، فسئل في هذا، فأجاب: لأني إذا صمت ضعفت عن الصلاة.

8 وكان ابن مسعود يقرأ القرآن، وشهد له الرسول بحسن القراءة، فقال فيه: "من سره أن يقرأ القرآن رطبا كما أنزل فليقرأه على قراءة ابن مسعود"، وهو أول من جهر بالقرآن من الصحابة، وذلك أنه عندما نزلت سورة الرحمان دعاه النبي إلى

Actions

Layer: Semantic Layer

Forward annotation?

Annotation

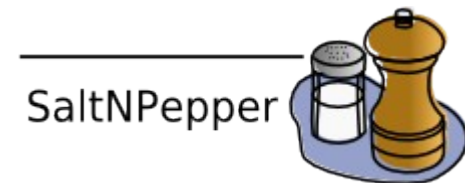
No annotation selected!

Technische Universität Darmstadt – Computer Science Department – WebAnno – 3.0.0-SNAPSHOT (2016-11-23 22:15:02, build a3582fdec85645161196d86fe23209c1619275a9)

# Search and visualization: Annis



- ✓ Extremely powerful searches
- ✓ Import flexibility: Pepper
- ✓ Portability



# Search and visualization: Annis

[Help/Examples](#)
[cobhuni.ocrd\\_texts](#)
[Frequency Analysis](#)
[Query Builder](#)
[Query Result](#)

[Help us to make ANNIS better!](#)

**tok & meta::source = "exegesis"**

Query Builder

Search More History

Valid query, click on "Search" to start searching.

Corpus List Search Options

Visible: All

Name	Texts	Tokens		
BerlinerSenatstexte_0.2	15	110,523	📘	📄
cobhuni.altafsir	2	2,164	📘	📄
<b>cobhuni.ocrd_texts</b>	<b>11</b>	<b>36,132</b>	<b>📘</b>	<b>📄</b>
complete.altafsir	18,530	12,601,880	📘	📄
complete.hadith_alislam	24,038	11,482,139	📘	📄
complete.ocrd_texts	13	37,755	📘	📄
testing_with_annotation	1	11	📘	📄

Filter documents by name

document name	corpus path	visualizer	Info
AlMulaqqin	cobhuni.ocrd_texts > AlMulaqqin	<a href="#">full text</a>	📘
AlTaufi	cobhuni.ocrd_texts > AlTaufi	<a href="#">full text</a>	📘
Fakihani	cobhuni.ocrd_texts > Fakihani	<a href="#">full text</a>	📘
Farhud	cobhuni.ocrd_texts > Farhud	<a href="#">full text</a>	📘
Fashni	cobhuni.ocrd_texts > Fashni	<a href="#">full text</a>	📘
Hajar	cobhuni.ocrd_texts > Hajar	<a href="#">full text</a>	📘
lyad	cobhuni.ocrd_texts > lyad	<a href="#">full text</a>	📘
Munawi	cobhuni.ocrd_texts > Munawi	<a href="#">full text</a>	📘
Nabrawi	cobhuni.ocrd_texts > Nabrawi	<a href="#">full text</a>	📘
Qurtubi	cobhuni.ocrd_texts > Qurtubi	<a href="#">full text</a>	📘
Rajab	cobhuni.ocrd_texts > Rajab	<a href="#">full text</a>	📘

# Search and visualization: Annis

Help us to make ANNIS better!

Base text v

1 < < 1 / 3 > > Displaying Results 1 - 10 of 28 Result for: /.\*ط?.?[[ق?[[س.\*/

1 Path: cobhuni.ocrd\_texts > AlMulaqqin (tokens 995 - 1005) left context: 5 right context: 5

وغير مخلقة § أي : السقط . وهذا التخليق والتصوير يكون

grid (default\_ns)

tok وغير مخلقة § أي : السقط . وهذا التخليق والتصوير يكون

tree (default\_ns)

2 Path: cobhuni.ocrd\_texts > AlTaufi (tokens 410 - 420) left context: 5 right context: 5

ذلك أنه لا يصلح على السقط حتى يستكمل أربعة أشهر وهي

grid (default\_ns)

tree (default\_ns)

3 Path: cobhuni.ocrd\_texts > Fakihani (tokens 568 - 578) left context: 5 right context: 5

: المصورة، وغير المخلقة : السقط؛ قاله ابن العالبة وغيره .

grid (default\_ns)

tree (default\_ns)

4 Path: cobhuni.ocrd\_texts > Fakihani (tokens 1131 - 1141) left context: 5 right context: 5

[ : لا أعتبر بإسقاط العلقه، وإنما الاعتبار بظهور الصورة

grid (default\_ns)

tree (default\_ns)

5 Path: cobhuni.ocrd\_texts > Fakihani (tokens 1158 - 1168) left context: 5 right context: 5

المتقدم [ ذكره ] وأن مسقطه العلقه أو المصغة يصدق على

grid (default\_ns)

tree (default\_ns)

6 Path: cobhuni.ocrd\_texts > Farhud (tokens 218 - 228) left context: 5 right context: 5

على أبي جهل بعد أن أسقط جريحا . كان ابن مسعود

grid (default\_ns)

tree (default\_ns)

7 Path: cobhuni.ocrd\_texts > Farhud (tokens 2582 - 2592) left context: 5 right context: 5

المصورة وغير المصورة وهذه تصير سقطا . وعن ابن مسعود أن

grid (default\_ns)

tree (default\_ns)

8 Path: cobhuni.ocrd\_texts > Hajar (tokens 3775 - 3785) left context: 5 right context: 5

...

Query Builder

Search

28 matches in 9 documents

Name	Texts	Tokens		
BerlinerSenatstexte_0.2	15	110,523		
cobhuni.altafsir	2	2,164		
<b>cobhuni.ocrd_texts</b>	<b>11</b>	<b>36,132</b>		
complete.altafsir	18,530	12,601,880		
complete.hadith_alislam	24,038	11,482,139		
complete.ocrd_texts	13	37,755		
testing_with_annotation	1	11		

# Search and visualization: Annis

◀ About ANNIS
Help us to make ANNIS better!

Help/Examples
cobhuni.ocred\_texts x
Frequency Analysis x
Query Builder x
Query Result x
cobhuni.ocred\_t... x

cobhuni.ocred\_texts > AIMulaqqin - Visualizer: full text

**tok & meta::source = "exegesis"**

Query Builder

Search More History

Valid query, click on "Search" to start searching.

Corpus List Search Options

Visible: All

Name	Texts	Tokens		
BerlinerSenatstexte_0_2	15	110,523		
cobhuni.altafisir	2	2,164		
cobhuni.ocred_texts	11	36,132		
complete.altafisir	18,530	12,601,880		
complete.hadith_alislam	24,038	11,482,139		
complete.ocred_texts	13	37,755		
testing_with_annotation	1	11		

عن أبي عبد الرحمن عبد الله بن مسعود قال حدثنا رسول الله صلى الله عليه وسلم وهو الصادق المصدوق: "إن أحذكم يجمع خلقه في بطن أمه أربعين يوماً نطفة، ثم يكون علقة مثل ذلك، ثم يكون مضغة مثل ذلك، ثم يرسل إليه الملك فينفخ فيه الروح، ويؤمر بأربع كلمات: يكتب رزقك، وأجله، وعمله، وشقي أو سعيد؛ فوالذي لا إله غيره، إن أحذكم يعمل بعمل أهل الجنة حتى ما يكون بيننا وبينها إلا ذراع، فيسبق عليه الكتاب فيعمل عمله كإيمان الذي عمله، وإن أحذكم يعمل بعمل أهل النار حتى ما يكون بينه وبينها إلا ذراع، فيسبق عليه الكتاب فيعمل عمله كإيمان الذي عمله، وشهد بدران والمشاهد كلها، وكان كثير الدخول عليه، مات بالكوفة - وقيل: بالمدينة - بعد الثلاثين؛ أما سنة اثنتين أو ثلاث. نانيها: في الفاظ ومعانيه، معنى "ثنا": أنشأ لنا خيراً حادثاً، وهو أصل فيما يستعمله المحذون من قولهم: "ثنا" من لفظ الشيخ، وإما قراءة عليه، وأبناً إجازة. و "الصادق": الآتي بالصدق، وهو الخبر المطابق، و "المصدوق": الذي يأتيه غيره بالصدق؛ فهو صادق في قوله وفيما يأتيه من الوحي، صادق فيما أخبر مصدوق فيما أخبر؛ لأن مصدوق أن الله صدقه فيما وعده به، وهذا تأكيد، وعلى هذا القياس الكاذب والمكذوب. ومنه: قول علي يوم نهروان: "والله ما كذبت ولا كذب من أخبرني" والشارح 200 جبريل يخبره، وعكسه ابن صباد حين قال: "بأنبي صادق وكاذب" و "أرى عرشاً على الماء" فهو إذاً كاذب مكذوب. ومعنى "يجمع": يضم ويحفظ مادة خلقه، وهو الماء الذي علق منه، و "العلقة": قطعة دم قبل أن تبس، و "المضغة": قطعة لحم قدر ما تمضغ؛ كعرقه بمقدار ما يغرف. و "الروح": هو المعنى الذي يحيى به الإنسان، وهو من أمر الله - تعالى - كما أخبر، وللناس في تحقيقه اختلاف كبير، ولفظه مشترك بين معاني، و "الرزق": ما يتناوله الإنسان في إقامة مدته من مأكول ومشروب وملبوس وغير ذلك، و "الأجل": مدة الحياة. ثالثها: قوله: "يكتب" هو بالياء الموحدة، وهو بدل من "أربع" و "شقي أو سعيد" مرفوع خبر مبتدأ محذوف: أي: وهو شقي وسعيد. وفي "صحيح ابن حبان" من حديث أبي الدرداء مرفوعاً: "فرغ الله إلى كل عبد من خمس: من رزقه، وأجله، وعمله، وأثره، ومضجعه - يعني: قبره؛ فإنه مضجعه على الدوام - وما تدري نفس بأي أرض تموت" وفي "تجريد الصحاح" لرزين، من حديث سهل بن سعد مرفوعاً: "إذا وقعت النطفة في الرحم... الحديث، وفيه: "أذكر أم أنثى؟ أشقى أم سعيد؟ وما عمره؟ وما [ق/٢٨-١] رزقه؟ وما أثره؟ وما مصائبه؟ فيقول الله، ويكتب الملك؛ فإذا مات الجسد دفن من حيث أخذ ذلك التراب". والمراد بالذراع: تمثيل القرب. رابعها: لا التفات إلى ما حكى عن عمرو بن عبدي - لا وكان من زهاد القدرية - من إنكار الحديث: فهو أقل من هذا! خامسها: بين الخطيب الحافظ في كتاب "الفصل للوصل" أن أول الحديث إلى قوله: "وشقي أو سعيد" وما بعده كلام ابن مسعود، ثم برهن لذلك. سادسها: ظاهر الحديث أن أعمال الحسنات والسيئات أمارات وليست بموجبات، وأن العاقبة في ذلك للسابقة، سابعا: قوله: "في بطن أمه أربعين يوماً" يريد: نطفة، قال بعض العلماء: ولذلك جعل على المتوفى عنها: أربعة أشهر وعشراً؛ لأن الأربعة لا تعتبر الخلق، وعشر احتياط، وغيرها ثلاث حيض؛ لأن عليها رقبياً، وأبيح لها أن تترنن تغايب زوجها. وجاء تفسيره عن ابن مسعود أن النطفة إذا وقعت في الرحم وأراد الله خلق بشر فيها؛ طارت في بشرة المرأة تحت كل طفر وشعر، ثم يمكت أربعين ليلة، ثم ينزل دماً في الرحم؛ فذلك جمعها، والذي في الحديث الذي يجمع خلقه أربعين يوماً بخلاف تفسيره، أنه يجمع بعد الأربعين، وسيأتي إيضاحه. ثامنها: جاء في رواية بحذف: "وعمله" والمراد: بكتبان أربعة أشياء من حاله رزقه: قلة وكثرة وحلالاً وحراماً. ومن أي جهة هو، ونحو ذلك، وأجله: طولاً وقصراً، وعمله: صالحاً وطالحاً، وشقي في الآخرة أو سعيد. ويجوز أن يكون المراد: ذكر جملة ما يؤمران به، لا أن كل شئ ص يؤمر فيه بهؤلاء الأربعة، وقد أسلفنا رواية "وأثره"، ويكون ذلك على كل شخص. ناسعها: قد أسلفنا الكلام على معنى الجمع؛ أن المنى يقع في في الرحم حين انتزاعه بالقوة الشهوانية الدافعة متفرقاً، فيجمعه الله في محل الولادة من الرحم في هذه المدة - كما أسلفنا عن ابن مسعود - ثم يكون علقة في مثل ذلك، وذلك الأول إشارة إلى المحل الذي اجتمعت فيه النطفة الملك الموكل بالرحم فينفخ فيه الروح "كما قال في حديث أنس: "إن أوصارت علقة، وذلك الثاني إشارة إلى الزمان الذي هو الأربعون. وكذا القول [ق/٢٨-ب] في قوله: "ثم يكون مضغة مثل ذلك، ثم يرسل 200 المضغة، وليس الأمر كذلك؛ إنما ينفخ فيها بعد أن تتشكل تلك المضغة بشكل ابن آدم؛ أي تتصور بصورته، كنا قال الف - تعالى - فدوكل بالرحم ملكاً... وظاهر هذا السياق أن الملك عند مجيئه ينفخ الروح في 200 تعالى: فخلقنا المضغة عظماً فكسونا العظم لحماً" وكما ذكر في الآية الأخرى: من مضغة مخلقة أي: مصورة غير مخلقة أي: السقط. وهذا التخليق والتصوير يكون في مدة أربعين يوماً، وحينئذ ينفخ فيه الروح، وهو المعنى بقوله تعالى: "ثم أنشأناه خلقاً آخر". العاشر: قدر النفخ هنا بعد مائة وعشرين يوماً، وصح في حديث آخر: بعد أربعين أو اثنين وأربعين يوماً، فيجمع بينهما بأن ذلك راجع إلى اختلاف الأجنة، أو بأن الملك ملازمه، ومرعاة لحال النطفة من الأربعين إلى تمام المائة والعشرين. وقد أوضحت الكلام عليه في "شرح صحيح البخاري" فراجعه منه تجد ما ينشفي العليل. وإدعى القاضي عياض أنه لم يختلف في أن نفخ الروح فيه إنما يكون بعد مائة وعشرين يوماً، وذلك تمام أربعة أشهر ودخوله في الخامس، وهذا موجود بالمشاهدة، وعليه يعول فيما يحتاج إليه من الأحكام والاستلحاق عند التنارع، ووجوب النفقات على حمل المطلقات، وذلك لتيقنه بحركة الجنين في الجوف. وقد قيل: إنه الحكمة لي عدة المرأة من الوفاة بأربعة أشهر وعشر - كما مر - وبعد الدخول في الخامسة تحقق براءة الرحم ببلوغ هذه المدة إذا لم يظهر حمل، ونفخت الملك في الصورة سبب يخلق الله فيها عنده الروح والحياة؛ لأن النفخ المتعارف إنما هو إخراج ریح من الناقع يتصل بالمنفوخ فيه، ولا يلزم منه عقلاً ولا عادة في حقنا [ق/٢٩-أ] تأثير المنفوخ فيه، وإن قدر حدوث شيء عند ذلك النفخ؛ فذلك باحداث الله لا بالنفخ، وغاية النفخ أن يكون مغدياً عادياً لا موجياً عقلياً، وكذلك القول في الأسباب المعتادة؛ فليتنامل هذا الأصل و يتمسك به لينجد من مذاهب أهل الضلال. وقد ظهر شر هذا الترتيب، وإن كانت القدرة صالحة لإيجاده وجميع المخلوقات في أسرع لحظة وأوحى أي: لأنه كان كذلك في سابق علمه. الحادي عشر: ظاهر قوله: "ويؤمر بأربع كلمات" الأمر بكتابتها ابتداءً، والمراد أن يؤمر بذلك بعد أن يسأل عنه فيقول: يا رب، ما الرزق؟ ما الأجل؟ ما العمل؟ وشقي أو سعيد؟... كما تضمنته الأحاديث المذكورة مع هذا الحديث. وفي "الصحيح" من طريق ابن مسعود وابن عمر: "أن النطفة إذا استقرت في الرحم أخذها الملك، فقال: أي رب، ذكر أم أنثى؟ شقي أم سعيد؟ ما الأجل؟ ما الأثر؟ بأي أرض تموت؟ فيقال له: انطلق إلى أم الكتاب؛ فإنك تجد قصة هذه النطفة، فينطلق فيجد قصتها في أم الكتاب فيلحق، فتأكل رزقها وتطأ أثرها؛ فإذا جاء أجلها قيضت، فدفنت في المكان الذي قدر لها" زاد في رواية من حديث ابن مسعود رضي الله عنه: "أن الملك يقول: يا رب، مخلقة أو غير مخلقة؟ فإن كان غير مخلقة قذفها الأرحام دماً، وإن قيل: مخلقة، قال: أي رب، ذكر أم أنثى؟... إلى آخر ما سلف. فالمراد بالاستقرار: صبرورة النطفة علقة ومضغة؛ لأن النطفة قبل ذلك غير مجتمعة - كما سلف - فإذا اجتمعت وصارت علقة أو مضغة أمكن حينئذ أن تؤخذ بالكف، وسماها: نطفة، في حال كونها علقة، أو مضغة باسم مئذنها. الثاني عشر: يستفاد مما ذكرنا أن المرأة إذا ألفت نطفة لم يتعلق بها حكم، بخلاف العلقة والمضغة، فإنه تنقض

# Search and visualization: Annis

Help/Examples | complete.altafsir | Query Builder | Query Result

Base text v

1 Path: complete.hadith\_alislam > hadith.al-islam-209-217\_39-3-17\_exegesis (tokens 115 - 125) left context: 5 right context: 5

اللَّهُ عَلَيْهِ وَسَلَّمَ - " اُخْفِصِي " ( فَقَدْ وَجِبَ الْغُسْلُ

2 Path: complete.hadith\_alislam > hadith.al-islam-9140-9140\_36-36-76\_exegesis (tokens 809 - 819) left context: 5 right context: 5

عَلَيْهِ وَسَلَّمَ يَا أُمَّ عَطِيَّةَ اُخْفِصِي وَلَا تَنْهَكِي فَإِنَّهُ أَنْصُرَ لِرُؤُوسِهِ

2 matches in 2 documents

Search More History

Corpus List Search Options

Visible: All

Name	Texts	Tokens		
BerlinerSenatstexte_0.2	15	110,523		
cobhuni.altafsir	2	2,164		
cobhuni.ocrd_texts	11	36,132		
complete.altafsir	18,530	12,601,880		
complete.hadith_alislam	24,038	11,482,139		
complete.ocrd_texts	13	37,755		
testing_with_annotation	1	11		

Info for salt:/complete.hadith\_alislam/hadith.al-islam-2... + x

Metadata

document: hadith.al-islam-209-217\_39-3-17\_exegesis

Name	Value
book	كِتَابُ الطَّهَّارَةِ
chapter	بَابُ وَاجِبِ الْغُسْلِ إِذَا اتَّقَى الْجَنَّاتَانَ
collection	الْمَنْتَقَى شَرْحَ مَوْطَأِ مَالِكٍ
href	http://hadith.al-islam.com/Page.aspx?pageid=192&BookID=39&PID=209
source	exegesis
annis:doc	hadith.al-islam-209-217_39-3-17_exegesis

corpus: complete.hadith\_alislam

# Search and visualization: Annis

The screenshot shows the Annis search interface. On the left, there is a search input field containing 'tok' and a search button. Below it, a table lists search results for 'tok' in the 'cobhuni.ocred\_texts' corpus. The table has columns for Name, Texts, and Tokens. The 'cobhuni.ocred\_texts' entry is highlighted. On the right, the search results are displayed in a list format, showing the path to the document and the text snippet. A modal window titled 'Query reference link' is open, displaying the URL for the search results and instructions on how to share the query.

Search results table:

Name	Texts	Tokens
BerlinerSenatstexte_0.2	15	110,523
cobhuni.altafsir	2	2,164
<b>cobhuni.ocred_texts</b>	<b>11</b>	<b>36,132</b>
complete.altafsir	18,530	12,601,880
complete.hadith_alislam	24,038	11,482,139
complete.ocred_texts	13	37,755
testing_with_annotation	1	11

Query reference link:

Share your query: 1. Copy the generated link  
2. Share this link with your peers.

[http://localhost:8080/annis-gui/#\\_q=dG9r&\\_c=Y29iaHVuaS5vY3JlZF90ZXh0cw&cl=5&cr=5&s=0&l=10](http://localhost:8080/annis-gui/#_q=dG9r&_c=Y29iaHVuaS5vY3JlZF90ZXh0cw&cl=5&cr=5&s=0&l=10)

# Search and visualization: Annis

⌵ About ANNIS
[Help us to make ANNIS better!](#)

tok

Query Builder

36132 matches  
in 11 documents

Visible:

Filter

Name	Texts	Tokens		
BerlinerSenatstexte_0.2	15	110,523		
cobhuni.altafsir	2	2,164		
<b>cobhuni.ocrd_texts</b>	<b>11</b>	<b>36,132</b>		
complete.altafsir	18,530	12,601,880		
complete.hadith_alislam	24,038	11,482,139		
complete.ocrd_texts	13	37,755		
testing_with_annotation	1	11		

Help/Examples
cobhuni.ocrd\_texts
Frequency Analysis
Query Builder
cobhuni.ocrd\_t...
Query Result

Showing histogram of top 100 results, see table below for complete dataset.

linear scale  
 log<sub>10</sub> scale

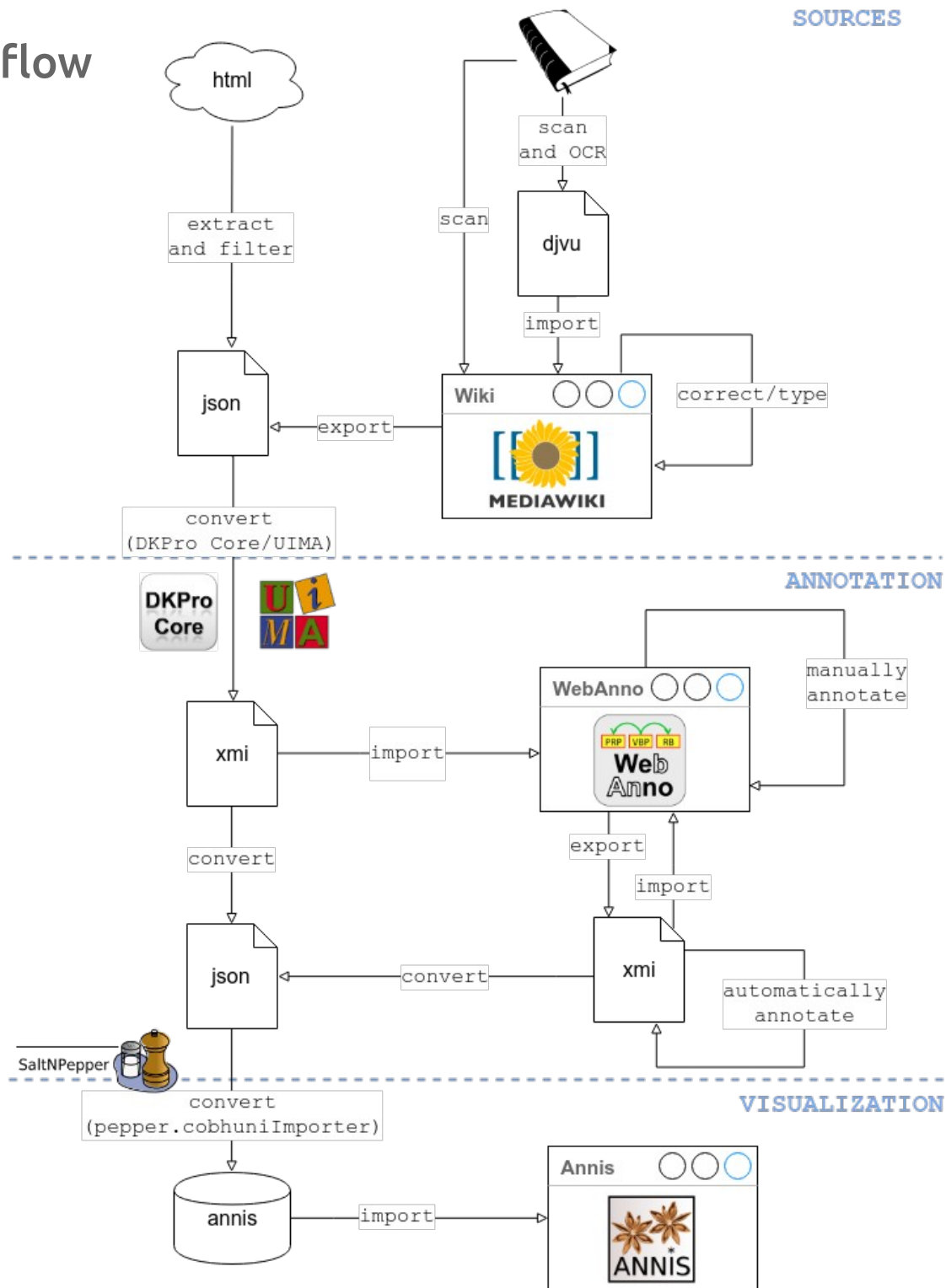
8038 items with a total sum of 36127 (query on cobhuni.ocrd\_texts)

rank	#1   spanned text	count
1	في	1,139
2	:	790
3	من	670
4	.	546
5	الله	470
6	أن	425
7	على	408
8	"	355
9	ما	338
10	ثم	327



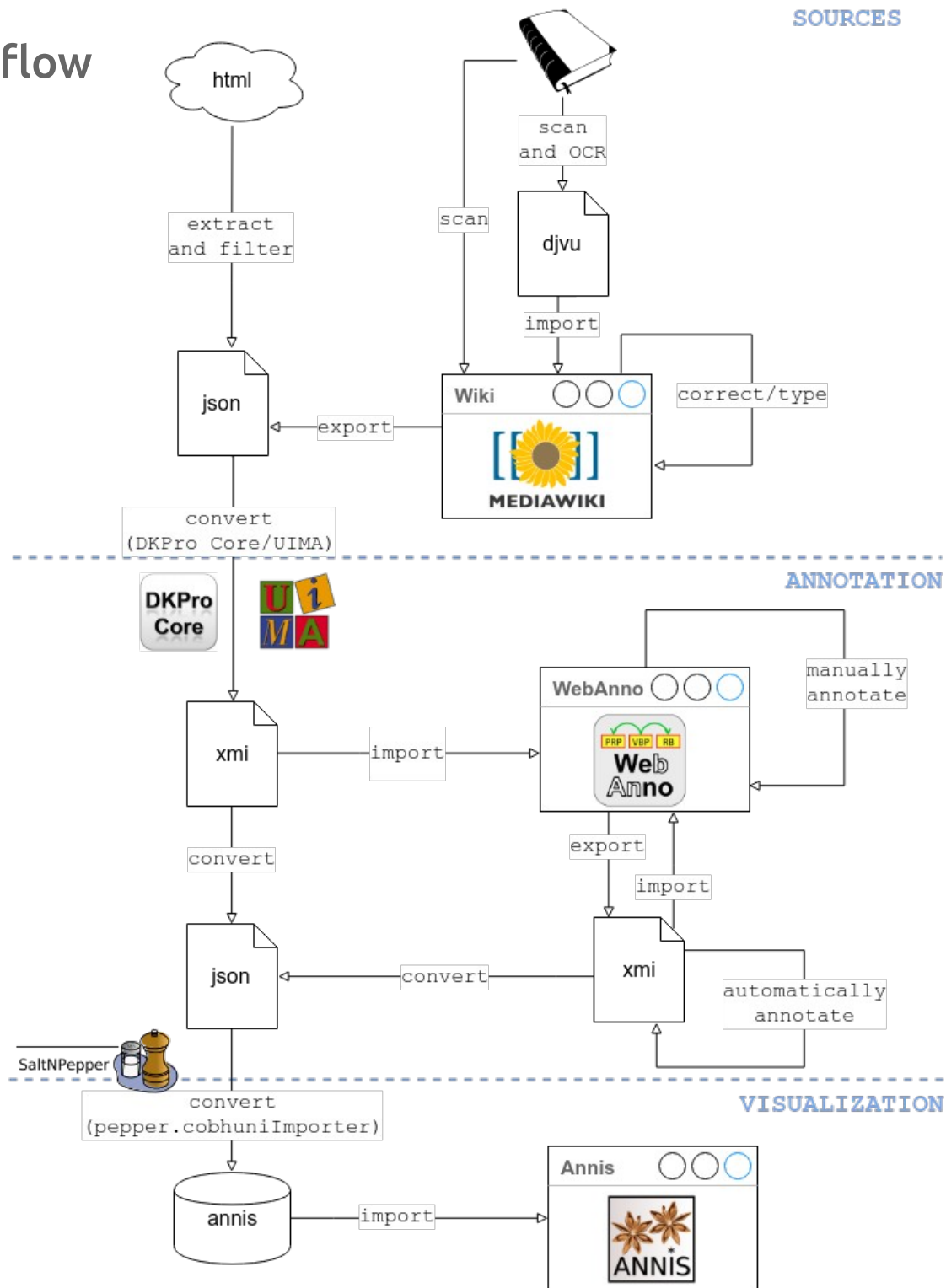
# Technical workflow

## SOURCES



# Technical workflow

## SOURCES



# Future work

- ✓ Continue annotating the corpus
- ✓ Include module for automatic annotation
- ✓ Add more text material (?)
- ✓ Improve the query capabilities: work on linguistic tokenization and morphological analysis

# Danke!