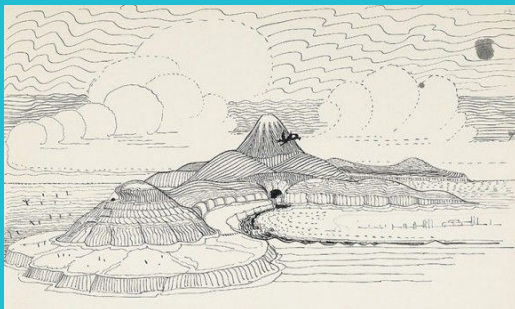


/instituut voor  
de Nederlandse  
taal/

## **The Gathering of the Clouds. Lexical resources & Linked Open Data**



Katrien Depuydt, Jesse de Does  
[katrien.depuydt@ivdnt.org](mailto:katrien.depuydt@ivdnt.org)  
[jesse.dedoes@ivdnt.org](mailto:jesse.dedoes@ivdnt.org)

# Intro INL lexica

- Dictionaries:
  - Modern Dutch: [anw.inl.nl](http://anw.inl.nl)
  - Historical Dutch: [gtb.inl.nl](http://gtb.inl.nl)
- GiGaNT-project
  - Molex: modern lexicon component
  - Hilex: historical lexicon component
  - modular; interlinked
- DiaMaNT-project (diachronic semantic lexicon)

# Modern lexicon (GiGaNT-Molex)

- Full paradigm and hyphenation information for 180000+ lemmata
- Continuous update with new words
- Published in lexicon service and (partly) in woordenlijst.org



gtb.inl.nl

2



# GiGaNT

/instituut voor de Nederlandse taal/

# GiGaNT-Hilex

IMPACT Attestation Tool - Mozilla Firefox

← → ↗ A impactdb.inl.loc/Projecten/Impact/Attestations/AttestationTool-new/attestationTool.php ☆ ↻

You have revised 6 lemmata (224 citations) so far. Last edited 2009-02-02 15:21:10 3910 lemmata in all. 10 to be revised yet...

Previous Next New Close 🚗 🔄 Before After YYYY-MM-DD hh:mm:ss

🔍 Lemma: **OOLIKHEID** >

- ⓧ ? Onder andere **oolikheên**, daar vervalschers gebruik van maken om tot hun schendig voornemen te raken, is 'er dit eene: zy leggen een beolyt papier op iemants schrift, en trekken de letters zoo na, +
- ⓧ ? In voegen dat 'er, in die manier van spreken, een **oolikheid**, en een arghjen is verborgen, +
- ⓧ ? Misschien is dit een **oolikheid** der Vorsten geweest, +
- ⓧ ? Met eene **oolikheid** heur slinkehand onder haare rokjes verbergende, zong zij enz. +
- ⓧ ? Schuldellooze en schalke **oolikheid**, +
- ⓧ ? 't Waar een' fraaye temper van Landvooghdye geweest, zoo men, met uitschot der **oolykheeden**, de pure deughden van hun beide (*Requesens en Alva*) hadde kunnen te zaamen mengen, +
- ⓧ ? 'T is klaar van **oolikheid** dat hij zoo lange slaapt (*en niet uit behoefte aan rust*). +

# DiaMaNT: diachronic semantic lexicon

lemma_id	mnw_lemma	mnw_mdl	pos	gram	syn	def
01449	ALWARICH	alwarig			Dom	<i>Dom, onnoozel.</i>
04666	BLINT	blind	bnw	bnw	Dom	<i>Dom, dwaas, onverstandig</i>
04739	BLONC	blonk	bnw	bnw	dom	<i>Overdrachtelijk. Stomp, pl</i>
05047	BOOS	boos	bnw	bnw	Dom	<i>Dom, onnoozel (synon. va</i>
05629	BRUUT	bruut	znw	znw(m.)	Dom	<i>Dom, onnoozel, en als znw bloed.</i>
06883	DOEM	dom	znw	znw(m.)	Dom	<i>Dom, domkerk, bisschops</i>
07004	DOMP	dom	bnw	bnw	dom	<i>Onontwikkeld, onverstandig, beperkt verstand of geest dwaas.</i>
07153	DORACHTICH	doorachtig	bnw	bnw	dom	<i>Dwaas, zot, dom, onverstandig, doregteg, idiota.</i>
07460	DORPERLIJIC	dorperlijk	bnw	bnw	dom	<i>Ook schijnt dorperlijk in d te staan.</i>
07941	DUL	dol	bnw	bnw	dom	<i>Dol, dom, dwaas, krankzinnig, onnoozel, onverstandig; v redeloos.</i>
08063	DUUSTER	duister	bnw	bnw	Dom	<i>Dom, onnoozel, niet door-</i>
08114	DWAES	dwaas	bnw	bnw	Dom	<i>Dom, bekrompen van vers stultus.</i>
09260	EELSC	ezels	bnw	bnw	dom	<i>Ezelachtig, dom.</i>
18841	IDIOOT	idiot	znw	znw(m.)	Dom	<i>Dom, onontwikkeld. Het w beteekent een bijzonder p. particulier, van gr. idiotas, ( vanwaar ook idioticon, w. particularismen of provinc eigenaardige uitdrukking), is datgene, waarmede het gebruikt wordt (niet-dokte dus leek) heeft in het mnl. wijs, niet geleerd, doch na tgw. betekenis onwijs, sin illiteratus, imperitus.</i>
21604	KINT	kind	bnw	bnw	dom	<i>Jong van jaren, kinderlijk kinderachtig, onontwikkeld</i>
21610	KINTSC	kinds	bnw	bnw	dom	<i>Als een kind handelende, i onnoozel, dom, dwaas. Vg 164 kintsch, puerilis. Kil. 1 kindisch, puerilis et infans inconsideratus instar pueri ndl. kindsch heeft zich ver</i>

lemma_id	wnt_lemma	wnt_mdl	pos	gram	syn	def
A000968	BÊTE	bête	bnw	bnw	dom	<i>Dom, onnoozel.</i>
A000968	BÊTE	bête	bnw	bnw	dom	<i>Dom, onnoozel.</i>
A008890	IGNORANT	ignorant	bnw	znw	dom	<i>Onwetend; dom. In de 16de e. vooral aangetroffen</i>
A014467	STUPIDE	stupide	bnw	bnw	dom	<i>Dom, stompzinnig.</i>
M009814	BOERSCH	boers	bnw	bnw	dom	<i>Oneigenlijk, in ongunstigen zin. Blijk gevende va</i>
M010624	BOT	bot	bnw	bnw	dom	<i>Van personen. Dom, stomp, onbevattelijk; het teg</i>
M010624	BOT	bot	bnw	bnw	dom	<i>Van personen. Dom, stomp, onbevattelijk; het teg</i>
M013932	DOMMELIJK	dommelijk	bnw	bnw	dom	<i>Bnw. — Min of meer dom, dom, suffig. In dit gel</i>
M017888	GEESTELOOS	geesteloos	bnw	bnw	dom	<i>Van personen. Van geest of verstand verstoken; be</i>
M022030	GROEN	groen	bnw	bnw	dom	<i>Met betrekking tot iemands kennis of verstand. Zc</i>
M022030	GROEN	groen	bnw	bnw	dom	<i>Van de om hun plomphheid van manieren en versta</i>
M022113	GROF	grof	bnw	bnw	dom	<i>Met betrekking tot het verstand. Bot, dom, onver</i>
M037638	LOMP	lomp	bnw	bnw	dom	<i>Van 's mensen uitingen, handelingen enz. Dom.</i>
M037638	LOMP	lomp	bnw	bnw	dom	<i>Van 's mensen uitingen, handelingen enz. Dom.</i>
M042799	ONBEDUIDEND	onbeduidend	bnw	bnw	dom	<i>Van iemands gelaat. Niet veel beduidend, weinig t</i>
M042906	ONBENULLIG	onbenullig	bnw	bnw	dom	<i>Bnw. Van personen. Geen benul hebbende, zonder</i>
M042906	ONBENULLIG	onbenullig	bnw	bnw	dom	<i>Bijw. — Op eene wijze die van geen benul of begrip</i>
M043110	ONBEZIND	onbezind	bnw	bnw	dom	<i>Eigenlijk, van menschen. Oververstandig, dom, dyc</i>
M043110	ONBEZIND	onbezind	bnw	bnw	dom	<i>Overdrachtelijk, van zaken, t.w. die een persoon b</i>
M047006	ONVERSTANDEL	onverstandel	bnw	bnw	dom	<i>Dwaas, dom; thans verouderd.</i>
M047671	OOLIK	olijk	bnw	bnw	dom	<i>Dom, dwaas, onnoozel.</i>
M047671	OOLIK	olijk	bnw	bnw	dom	<i>Dom, dwaas, onnoozel.</i>
M054328	PLAT	plat	bnw	bnw	dom	<i>Eenvoudig, dom, onnoozel, sullig. Thans veroude</i>
M054328	PLAT	plat	bnw	bnw	dom	<i>Van handelingen. Onnoozel, dom.</i>
M061360	RUID	ruid	bnw	bnw	dom	<i>Met betrekking tot personen. Onbeschaafd, onont</i>
M061368	RUIDIG	ruidig	bnw	bnw	dom	<i>Onbeschaafd, onontwikkeld, ruw, grof, plomp, bo</i>
M063919	SLECHT	slecht	bnw	bnw	dom	<i>Onnoozel, onwetend, dom.</i>
M065730	STAAS	staas	bnw	bnw	dom	<i>In W.-Vlaanderen. Dwaas, dom, bot. Verg. STAAS</i>
M071944	UITHEEMSCH	uitheems	bnw	bnw	dom	<i>Dom, onwetend. W.g. en veroud.</i>
M071944	UITHEEMSCH	uitheems	bnw	bnw	dom	<i>Dom, onwetend. W.g. en veroud.</i>
M072530	UITZINNIG	uitzinnig	bnw	bnw	dom	<i>In verzwakte bet.: blijkgevend van een gemis aan ;</i>
M075227	VERDORVEN	verdorven	bnw	bnw	dom	<i>In zwakkere bet.: dom, kortzichtig.</i>
M078479	VERSTANDELOOS	verstandeloos	bnw	bnw	dom	<i>Zonder verstand ; dom; bot, lomp, stompzinnig.</i>
M079269	VERWATEN	vervaten	bnw	bnw	dom	<i>Niet over verstand beschikkend, redeloos; van het</i>
M082952	VROOM	vroom	bnw	bnw	dom	<i>Ook: onnoozel, dom.</i>
M085928	WIEDES	wiedes	bnw	bnw	dom	<i>Dom, onnoozel, simpel: onkundig.</i>
M085928	WIEDES	wiedes	bnw	bnw	dom	<i>Dom, onnoozel, simpel: onkundig.</i>
M089793	ZOT	zot	bnw	bnw	dom	<i>In het Dict. Tetragl., mog. ook van uitingen: flauw</i>
M089793	ZOT	zot	bnw	bnw	dom	<i>Van personen: niet in het bezit van gezond verstan</i>
M089793	ZOT	zot	bnw	bnw	dom	<i>Van eigenschappen, gevoelens, handelingen e.d.: v</i>
M089793	ZOT	zot	bnw	bnw	dom	<i>Inz. van uitingen in woord en geschrift: zonder zi</i>



# Ways of using using the lexica

- Just manual lookup
- Use in computational tools (Nederlab tagging, INL-labs, ...)
- Use in search scenario's
  - Simple query expansion
  - More general: “chaining search”

# Lexicon service

## From wordform to lemma:

[http://sk.taalbanknederlands.inl.nl/LexiconService/lexicon/get\\_lemma?database=lexicon\\_service\\_db&wordform=loopt](http://sk.taalbanknederlands.inl.nl/LexiconService/lexicon/get_lemma?database=lexicon_service_db&wordform=loopt)

From lemma to wordforms:

[http://sk.taalbanknederlands.inl.nl/LexiconService/lexicon/get\\_wordforms?database=lexicon\\_service\\_db&lemma=lopen](http://sk.taalbanknederlands.inl.nl/LexiconService/lexicon/get_wordforms?database=lexicon_service_db&lemma=lopen)

```
▼<root>
  ▼<lemmata_list>
    ▼<found_lemmata>
      <lemma>lopen</lemma>
      <pos>VRB</pos>
    </found_lemmata>
    <query_pos/>
    <query_word>loopt</query_word>
  </lemmata_list>
  <message>OK</message>
</root>
```

```
<root>
  <message>OK</message>
  ▼<wordforms_list>
    <query_pos>/>
    <query_word>lopen</query_word>
    <found_wordforms>elopen</found_wordforms>
    <found_wordforms>gelopen</found_wordforms>
    <found_wordforms>ghelopen</found_wordforms>
    <found_wordforms>ghelopen</found_wordforms>
    <found_wordforms>laupe</found_wordforms>
    <found_wordforms>liep</found_wordforms>
    <found_wordforms>liepen</found_wordforms>
    <found_wordforms>liept</found_wordforms>
    <found_wordforms>loepen</found_wordforms>
    <found_wordforms>loop</found_wordforms>
    <found_wordforms>loope</found_wordforms>
    <found_wordforms>loopen</found_wordforms>
    <found_wordforms>lopend</found_wordforms>
    <found_wordforms>lopende</found_wordforms>
    <found_wordforms>lopenden</found_wordforms>
    <found_wordforms>loopens</found_wordforms>
    <found_wordforms>loopenen</found_wordforms>
    <found_wordforms>looppene</found_wordforms>
    <found_wordforms>lloopt</found_wordforms>
    <found_wordforms>lope</found_wordforms>
    <found_wordforms>lopen</found_wordforms>
    <found_wordforms>lopende</found_wordforms>
  </wordforms_list>
</root>
```



Boeken Basis

Tekst

wereld or waereld or waerelds or wareld or weer

Eenvoudig zoeken

Historische spelling

+ waerelden

+ waerelt

+ weëled

+ weirelts

+ wèreld

+ werelden

+ werelds

Periode

1486 t/m 2013

Titel

Auteur

Taal

Onderwerp

Drukker/Uitgever

Wissen



#### Periode

- ☐ 15e eeuw (1)
- ☐ 16e eeuw (31)
- ☐ 17e eeuw (239)
- ☐ 18e eeuw (5641)
- ☐ 19e eeuw (6055)
- ☐ 20e eeuw (22936)
- ☐ 21e eeuw (120)

Kies periode...

#### Herkomst

Kies herkomst...

**35.345 boeken** gevonden voor: wereld of waereld of waerelds of wareld of weereld of weerelden of werelds of weerelt of weireld of weirelt \ "1486" t/m "2013"

Resultaten per pagina 10

1 2 3 ... 3535

Ga naar pagina 1 ga

Sorteer op relevantie

Weergave



dbnl



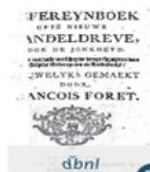
#### Refereynen en andere gedichten uit de XVIe eeuw



gebroken. Ten proposte van desen gesproken, wat is de weirelt & tvolck dat daer binnen // is? Al mach tvolk ...

Jaar van uitgave 1879-1881

[Meer details](#)



dbnl



#### Refereynboek ofte nieuwe wandeldreve, voor de jonkheyt



Refereyn van de Helle Wee my katyvig mensch! Wee dat ik ben geboren, Ter weireld oyt gebragt myn' ...

Auteur Foret, François

Jaar van uitgave [1667]

[Meer details](#)

zoeken in tekst regering

zoeken in titelgegevens 0-1900

zoeken in auteursgegevens

beperk samenstelling  
onzelfstandige titels  
zelfstandige titels  
koepeltitels  
herdrukken uitsluiten

beperk tot genre(s)

reset

zoek

- ☐ fictie
- ☐ non-fictie
- ☐ periodieken
- ☐ bloemlezing
- ☐ verzameld werk
- ☐ verzamelhandschrift
- ☐ hertaling

beperk tot collectie(s)

- CRM
- DBNL
- KB kranten
- EDBO
- HI-Acta Zuid-Holland
- HI-Dagboek de Clercq
- Staten-Generaal Digitaal
- SoNaR

aanwezigheid tekst

tekst aanwezig

verrijkingen aanwezig

woordvarianten

varianten inbegrepen

reset

zoek

- ☐ woordsoorten
- ☒ varianten
  - ☒ regeering (1057595)
  - ☒ regeeringe (16065)
  - ☒ regeeringen (31940)

alle woorden: "regering", gezocht in tekst, woordvarianten inbegrepen, metadata van bron: publicatieperiode: 0 - 1900, tekst aanwezig | 914.326 documenten gevonden

Reset alles

bewaren als corpus

lijst van bronnen | visueel overzicht | tijdlijn | statistieken | frequentielijsten | groeperingen

☒ titelmetadata ☐ lemma ☐ pos ☐ kenmerken ☐ entiteiten ☐ toon labels 5 linkercontext 5 rechtercontext

download kwic p. 1 van 36574

jaar van uitgave

1 2 3 4 5 6 7 8 9 > >>

## De natuurrkunde van het geheelal: Een 13de-eeuws middelnederlands leerdicht

datering: ca. 1275-1300

auteur: anoniem Natuurrkunde van het geheelal, De (-)

genre: fictie, non-fictie, poezie, leerdicht, artesliteratuur, natuurrkunde, sterrenkunde

collectie: DBNL

aantal hits: 3

tot aan het einde der regering van het huis van Burgondie  
tot aan het einde der regering van het huis van Burgondie  
Id . . Over de regeering der voorzetsels in het Middelnederlandsch

## Lof der Zotheid

datering: 1560

auteur: Johan Geillyaert (Brugge, ca. 1507-Emden, 1574) Desiderius Erasmus (Rotterdam (Zuid-Holland), 1466-Bazel, 1536)

genre: fictie, proza, poezie, satire, lofdicht(en)

collectie: DBNL

aantal hits: 6

behoort te onthouden van de regeringe der gemeente : hoewel hy  
hy vorderlick was met zijn regeringe . Want die aert der  
dat de gene die de regeeringhe heeft aenghenomen , niet zijn

Laad meer snippets

# Code sample

```
import scala.xml._

object LexiconService
{
  case class LexiconService(serviceURL: String, database: String)
  {
    def getWordforms(lemma: String):List[String] =
    {
      (XML.load(s"$serviceURL/get_wordforms?database=$database&lemma="
        + lemma.toLowerCase)
        \\ "found_wordforms").toList.map(_._text.toLowerCase).filter(_._matches("^(\\p{L})*$"))
    }

    def getLemmata(wordform: String):List[(String,String)] =
    {
      (XML.load(s"$serviceURL/get_lemma?database=$database&wordform=" + wordform.toLowerCase)
        \\ "found_lemmata").toList.map(e => ((e \ "lemma").text, (e \ "pos").text))
    }
  }

  val serviceURL = "http://sk.taalbanknederlands.inl.nl/LexiconService/lexicon"
  val database = "lexicon_service_db"

  def main(args: Array[String]) : Unit =
  {
    val ls = new LexiconService(serviceURL, database)
    ls.getWordforms("wereld").foreach(println)
    ls.getLemmata("ook").foreach(println)
  }
}
```

# Limitations

- Web services like this are simple to use but
  - each is different
  - *“This is all very nice, but can I just access your database?”*

So we need a more generic model to publish the data



# Clariah search

(in WP3, linguistics) search corpora, treebanks, lexica

- *local search* (ensure there are “endpoints”)
- *federated search* (jointly query similar datasources)
- *chaining search* (combine heterogeneous data sources)



# Clariah search

(in WP3, linguistics) search corpora, treebanks, lexica

- *corpora* (“*token-based annotation*”)
- *treebanks* (syntactic annotation)
- *lexica*
  - morphosyntactic
  - semantic
  - ...

# Clariah federated search

Multi-layer FCS query `(([pos="(AA|ADJ).*"]([pos="N.*"])([word="en"]([pos="(AA|ADJ).*"]([pos="N.*"])))2)`

Search for Any Language Multi-layer Federated Content Search Query Language (FCS-QL) in 5 selected collections

5 matching collections found

OpenSoNaR in blacklab server – Instituut voor de Nederlandse Taal

volksinitiatief zijnhier immers nog geen politieke realite  
ziek ofzo ) naja , zalige afmixing en verdere afwerking  
worden , doe daar een pittig rokje en hip kapsel e  
overal hetzelfde uitziet , met dezelfde winkels en dez  
zelfde dacht iedereen heeft een eige wil en eige m  
rode glansballen en piek , verder zilver en rode kra  
-morgan- overzeas ( zalige broeke en speciale t-shi  
overal hetzelfde uitziet , met dezelfde winkels en dez  
klant geworden doe daar een pittig rokje en hip ka

Corpus of Contemporary Dutch – Instituut voor de Nederlandse Taal

zaak maar zoethouden met allerlei prachtige beloften op  
massaal geschrokken van die harde fluitende wind en  
kan de verwoestende gevolgen van felle branden en  
motie misleidend is naar het Surinaamse volk en vol  
ouwels uit Turnhout Limburgs boekweitmeel Hasseltse  
aan elkaar van goedkope effecten roekeloze metaforen  
huidige wetgeving verbiedt voorlopig huisvesting militair  
in een psychedelische trip vol dynamische schokken e  
ter  
Ook voor de combitickets met Oude Kunst en Moderne

Letters as Loot corpus – Instituut voor de Nederlandse Taal

siekte en ik verbleef u toegeneegende vriend en toekomstige compeer en toekomstige Paater of u kent ik wense

Staten Generaal in Blacklab – Instituut voor de Nederlandse Taal

minachtend oog van allen die eene volksjustitie en eigen regt en nationale instellingen weten te waardenen blootstellen Neen  
be drijvigheid door zijn handel zijne fabrieken en zijne binuen en buitenlandse scheepvaart Ik neem de vrijheid  
deze Minister heeft aanvaard d.w.z. gelijke rechten en gelijke plichten en gelijke ontwikkelingskansen en bestaansmogelijkheden Meer mag m.  
behoud van de uitvoerrechten op ruwe gezouten en gedroogde huiden en eene heffing van de invoerrechten op het  
belangrijk zou zijn bij voorbeeld pharmaceutische chemie en chemische toxicologie en bijzondere beoefening van de kennis der medicinale

Collections

5 selected collections

Search for collection Select all Deselect all

- INT corpora – Homepage  
INT corpora  
Collapse (2 subcollections)
- INT historical Dutch corpora – Homepage  
INT historical Dutch corpora  
Collapse (4 subcollections)
- Letters as Loot corpus – Homepage  
Letters as Loot corpus
- Nederlab – Homepage  
Nederlab
- Staten Generaal in Blacklab – Homepage  
Staten Generaal in Blacklab
- The corpus Gysseling – Homepage  
The corpus Gysseling
- INT modern Dutch corpora – Homepage  
INT modern Dutch corpora  
Collapse (2 subcollections)
- Corpus of Contemporary Dutch – Homepage  
Corpus of Contemporary Dutch
- OpenSoNaR in blacklab server – Homepage  
OpenSoNaR in blacklab server
- The Språkbanken corpora – Homepage  
The Språkbanken corpora  
Expand (1 subcollections)

Close

<https://spraakbanken.gu.se/ws/fcs/2.0/aggregator/>

# Why linked open data??

- Choice as main interoperability standard for CLARIAH
- Interoperability without imposing a common schema on data
- Simple but expressive data model
- Aggregation and federation: easy transition
  - Compare with CLARIAH fcs 2.0 etc.
- Although
  - Unsuitable for content management (at least, for now, for us)

# Tower of babel problem

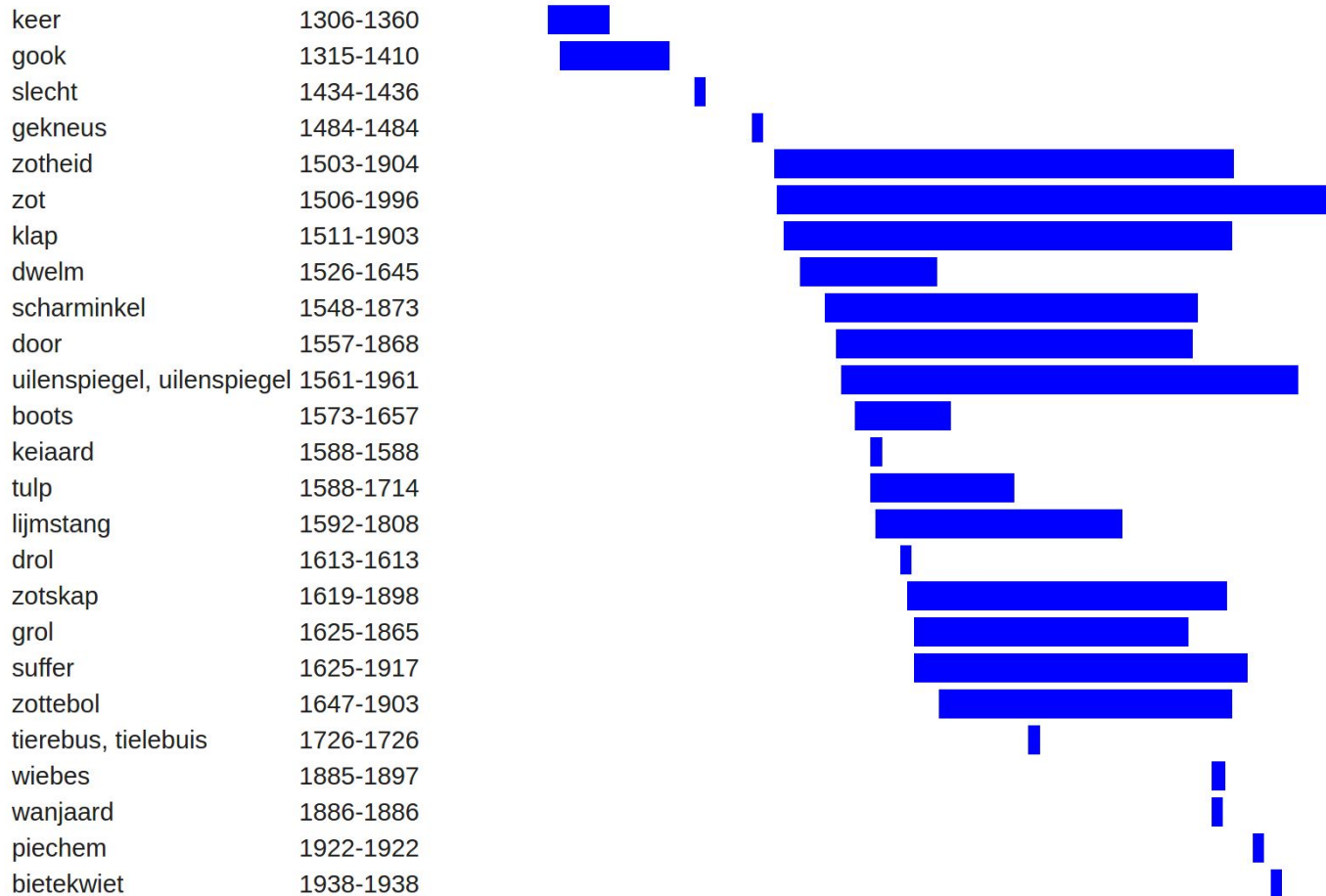
Each resource speaks its own language

- query language
- data encoding
- linked open data and standard encoding schemas as interlingua / glue
- universal dependencies

# DiaMaNT/LOD use cases ...

- Diachrony of similar terms
- Combining semantic lexica
- Using distributional information
- “Join” with external ontology data
- Deploy in corpus search

# Attestation-based chronology of related terms











# Use lexicon in corpus search

Per Hit

Per Document

Hits grouped

Documents grouped

Total hits: 132

Total pages: 7

Prev

1

2

3

4

5

6

7

Next

Toggle titles

Left context ▼	Hit text	Right context ▼	Lemma	Part of speech
... ridder sprac v tale es	sod	ende houerdeh vtermaten darombe moti ...	ZOT	ADJ(number=sg,inflection=0)
... so sere en wah hoe	dul	wasic dat ix v nine ...	DUL	ADJ(number=sg,inflection=0)
... van zuenkerke andie west side vanden	dullen	weghe tusschen weitins papen damme ...	DUL	ADJ(number=sg,inflection=n)
... liede bringhen vort dinken hem [wesen]	dulle	word ende tsine dinket hem ...	DUL	ADJ(number=pl,inflection=e)
... vier ander uolc es daer	onuroeder	die haren uader ende hare ...	ONVROED	ADJ(number=sg,inflection=0)
... houesce maniere na desen stommen	dullen	dieren alsi noten deen rijd ...	DUL	ADJ(number=pl,inflection=n)
... emele onder derde om hare	dorpere	houarde [in] [arderike] [balch] [hi] ...	DORPER	ADJ(number=sg,inflection=e)
... die gherualke bedieden wel die	dorpar	scalke [die] [hem] [niet] [ne] ...	DORPER	ADJ(number=pl,inflection=0)
... lijef ende wert sijn danden	dorperen	scuwt hi es lelic in ...	DORPER	ADJ(number=sg,inflection=n)
... uoglen dar vt somwile eist so	sot	dathet maket so uele sijn ...	ZOT	ADJ(number=sg,inflection=0)
... voeden ne connen no ouden	dorper	mensche merc hier up wel hier ...	DORPER	ADJ(number=sg,inflection=0)
... broet in der haghe die	dorpre	die lettelt doech hi broedet ...	DORPER	ADJ(number=sg,inflection=e)
... dan nes hi niet so	dul	alsem de buec idel es ...	DUL	ADJ(number=sg,inflection=0)
... slaet to doot om hare	dorper	gulseit groot ende dan verwecket ...	DORPER	ADJ(number=sg,inflection=0)
... uoeden met eweliken lone enten	dorpre	metten traghe sal hi in ...	DORPER	ADJ(number=sg,inflection=e)
... de ghene diene iaghen so	dul	es hi inder haghen alse ...	DUL	ADJ(number=sg,inflection=0)
... dier bediet al openbare den	ongheualleghen	sondare die wel beuroet ende ...	ONGEVALLIG	ADJ(number=sg,inflection=n)
... si comen dar si sijn	onuroeder	entan uerebben vp .i. sant ...	ONVROED	ADJ(number=sg,inflection=0)
... wel ter curen al eist	onwijs	in onsen sin wanen die ...	ONWIJS	ADJ(number=sg,inflection=0)
... want dus bliuen die wise	sot	ende die simple die wel ...	ZOT	ADJ(number=sg,inflection=0)

Prev

1

2

3

4

5

6

7

Next

Toggle titles

# Federated vs aggregated

Each resource speaks its own language

- convert or
- map

# Chaining search: corpora and lexica

```
SELECT distinct ?pw ?ow ?lvw ?lvl
```

```
WHERE {
```

```
  SERVICE lassyklein:
```

```
  {
```

```
    ?pred ud:nsbj ?onderwerp .
```

```
    ?pred ud:obj ?lv .
```

```
    ?onderwerp :word ?ow .
```

```
    ?pred :word ?pw .
```

```
    ?lv :word ?lvw
```

```
  } .
```

```
graph <http://rdf.ivdnt.org/lexica/dbnl_similarity/>
```

```
{
```

```
  ?sim diamant:source "woede" . ?sim diamant:target ?lvl
```

```
} .
```

```
graph <http://rdf.ivdnt.org/lexica/diamant/v1.0/>
```

```
{
```

```
  ?e ontalex:canonicalForm ?lvc .
```

```
  ?lvc ontalex:writtenRep ?lvl .
```

```
  ?e ontalex:lexicalForm ?lvf .
```

```
  ?lvf ontalex:writtenRep ?lvw
```

```
}
```

```
}
```

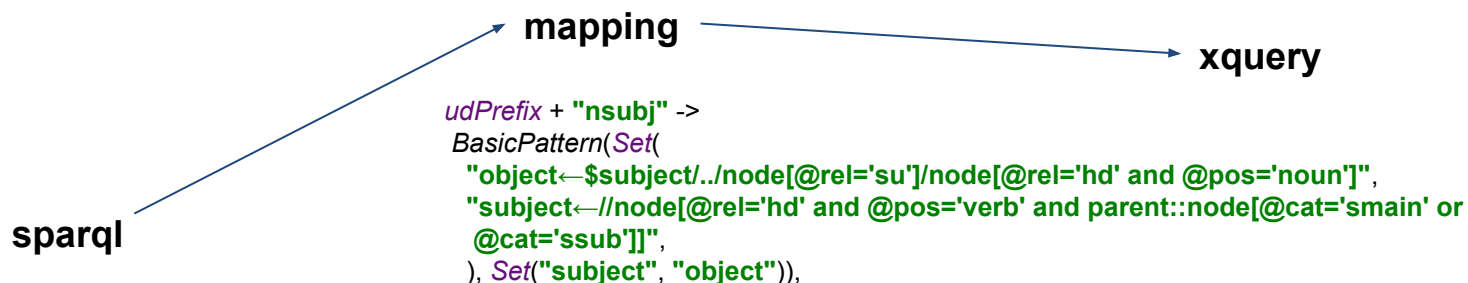
Triples predicate, object, subject from corpus

Distributional similarity to target word

Word forms from lexicon



# Chaining search: including treebanks



**prefix** : <http://example.org/>

**prefix** ud: <http://universaldependencies.org/u/dep/>

**select** ?tonderwerp ?tgezegde ?tlv ?sent **where**

```
{
  ?gezegde ud:nsubj ?onderwerp .
  ?gezegde ud:obj ?lv .
  ?lv :precedes ?onderwerp .
  ?gezegde :text ?tgezegde .
  ?onderwerp :text ?tonderwerp .
  ?lv :text ?tlv .
  ?gezegde :sentence ?sent
}
```

```
for $gezegde in (//node[@rel='hd' and @pos='verb' and
  parent::node[@cat='smain' or @cat='ssub']]),
$lv in ($gezegde/..node[@rel='obj1']/node[@rel='hd' and
  @pos='noun']),
$onderwerp in ($lv/ancestor::node/descendant::node[xs:int(@begin) >
  xs:int($lv/@begin)]) intersect
($gezegde/..node[@rel='su']/node[@rel='hd' and @pos='noun']),
$tgezegde in (string-join($gezegde//@word, ' ')),
$tonderwerp in (string-join($onderwerp//@word, ' ')),
$sent in ($gezegde/ancestor::alpino_ds/sentence),
$tlv in (string-join($lv//@word, ' '))
return( <result><var><name>tlv</name><value>{$tlv}</value></var>
<var><name>tgezegde</name><value>{$tgezegde}</value></var>
<var><name>sent</name><value>{$sent}</value></var>
<var><name>tonderwerp</name><value>{$tonderwerp}</value></var>
</result>)
```

# Discussion: future of tool use in humanities

- what can we expect non-technical researchers to do
- which tools are missing most
- or need more work to be usable