

The Cornetto Database



RU | STEVIN

Overview

1. Project background information
2. Alignment of lexical resources
3. Database design
4. Usage of the database in NLP

Goal

- <http://www.let.vu.nl/onderzoek/projectsites/cornetto/start.htm>
- Stevin tender project to develop a lexical semantic database for Dutch:
 - 40K Entries
 - Generic and central part of the language
- Data:
 - Combination of WordNet and FrameNet
 - Vertical and horizontal semantic relations
 - Combinatorial lexical constraints
 - Aligned with the English Wordnet
 - Extended with an ontology
- Automatic acquisition toolkit
- Started April 2006, ends March 2008
- Licensed from TST-centrale, Nederlandse Taalunie

Consortium

- Vrije Universiteit Amsterdam, Faculteit der Letteren, General Linguistics Department, Onderzoeksgroep Lexicologie/Terminologie:
 - Willy Martin, Isa Maks, Hennie vd Vliet, Roxane Segers, Piek Vossen
- Universiteit van Amsterdam, Instituut voor Informatica:
 - Maarten de Rijke, Erik Tjong Kim Sang, Katja Hofmann
- Katholieke Universiteit Leuven, Interdisciplinair Centrum voor Recht en Informatica (ICRI):
 - Sien Moens, Jan de Beer
- Irion Technologies BV:
 - Joop van Gent, Hetty van Zutphen, Piek Vossen

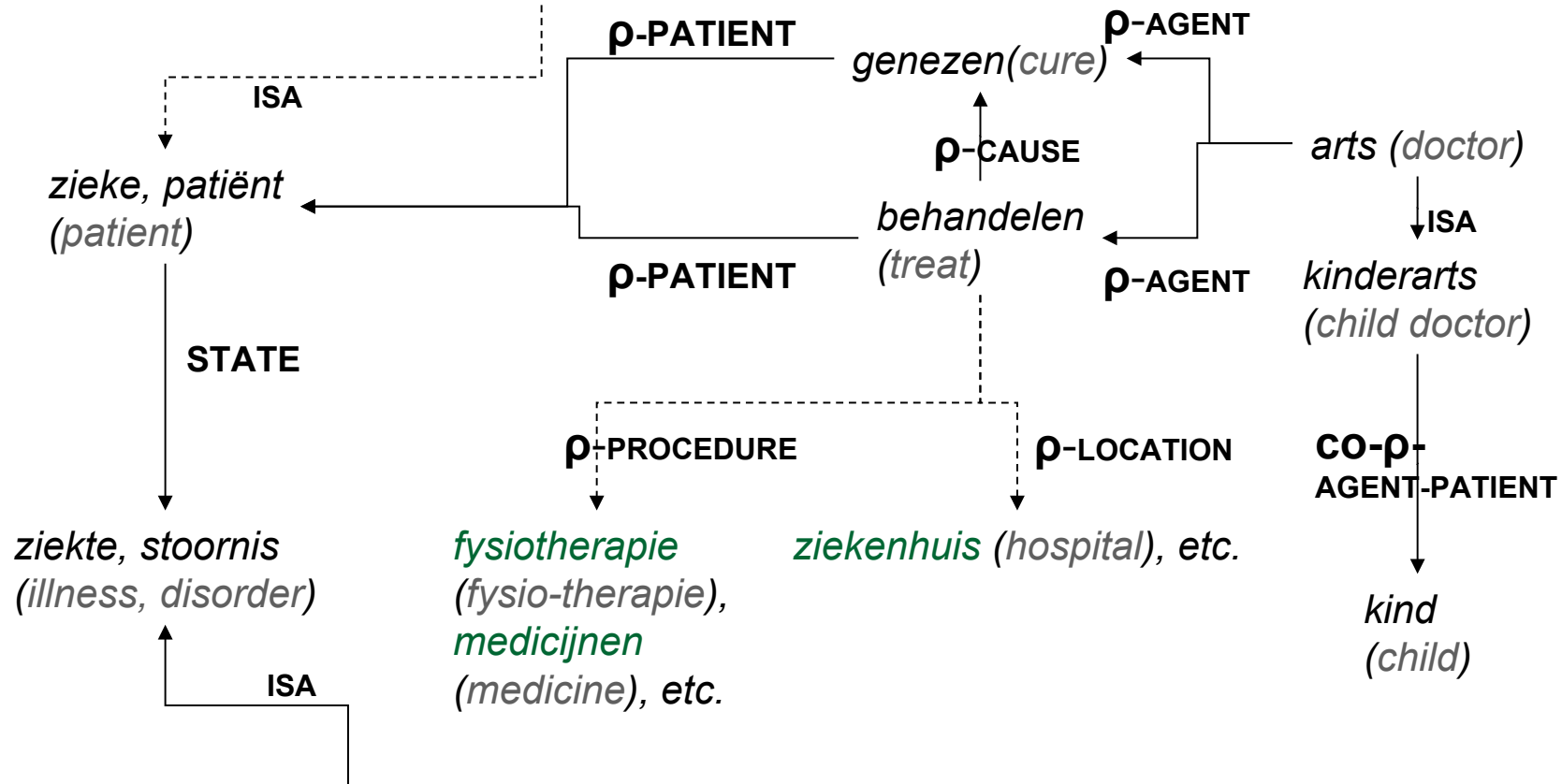
Other partners

- User-group:
 - ❑ Polderland
 - ❑ Knowledge Concepts
 - ❑ LibRT
 - ❑ Irion Technologies
 - ❑ Van Dale Lexicografie
 - ❑ Larcier-De Boeck
 - ❑ Rik Schutz
 - ❑ Gridline

- Ontology-group:
 - ❑ Dr. W. Ceusters, Office Line Engineering nv
 - ❑ Prof. F. van Harmelen, Vrije Universiteit Amsterdam
 - ❑ Dr. P. Buitelaar, DFKI
 - ❑ Dr. P. Monachesi, Universiteit van Utrecht

Horizontal & vertical semantic relations

chronisch zieke (chronical patient), *langdurig zieke* (long-term patient),
psychisch/geestelijk zieke (mental patient)



maagaandoening (stomach disorder)
nieraandoening (kidney disorder), *keelpijn* (sour throat).

Combinatorics

slots	fillers (lex/conc)	fillers (coll)
action	<i>behandelen</i> (someone treat)	<i>iem. behandelen</i>
theme	<i>patiënt</i> (a patient treat)	<i>een patiënt behandelen</i>
state	<i>ziekte</i> (someone treat for a disease)	<i>iem. behandelen <u>voor</u> een ziekte</i> <i>iem. <u>aan</u> zijn verwondingen</i> <i>(somene at his injuries treat)</i> <i>een ziekte behandelen (a disease treat)</i>

Approach

- Combine the information from two existing Dutch lexical resources:
 - The Dutch wordnet: synsets and lexical semantic relations
 - The Referentiebestand Nederlands: morpho-syntactic information, semantic information, pragmatic information, frame structures, lexical functions and combinatorics
- Macro level alignment
- Micro level alignment
- Populate with an ontology



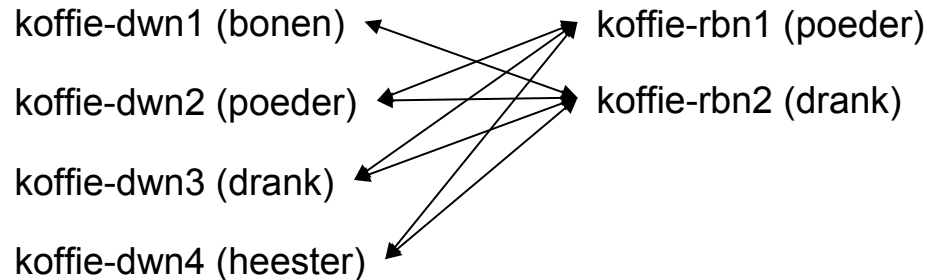
Global planning

- Two year project:
 - Month 1-6: design and database
 - Month 1-6: automatically aligned data
 - Month 7-10: ontology assignment
 - Month 7-22: editing
 - Month 7-15: acquisition
 - Month 16-17, 23-24: task-based evaluation

Alignment of lexical resources

Alignment

- Generate all weighted combinations:



- Produce merged output with mappings above probability threshold:

- New structure of word meanings

- koffie-cbn1(bonen) (source dwn1)
- koffie-cbn2 (poeder) (source dwn2, rbn1)
- koffie-cbn3 (drank) (source dwn3, rbn2)
- koffie-cbn4 (heester) (source dwn4)

Strategies for the macro-alignment

	Conf.	Dev.	Factor	LINKS	
1: 1 RBN & 1 DWN meaning, no synonyms	97.1	4,9	3	9936	8,1%
2: 1 RBN & 1 DWN meaning	88.5	8,6	3	25366	20,8%
3: 1 RBN & >1 DWN meaning	53.9	8,1	1	22892	18,7%
4: >1 RBN & 1 DWN meaning	68.2	17,2	1	1357	1,1%
5: overlapping hyperonym word	85.3	23,3	2	7305	6,0%
6: overlapping hyponyms	74.6	22,1	2	21691	17,7%
7: overlapping domain-clusters	70.2	15,5	2	11008	9,0%
8: overlapping definition words	91.6	7,8	3	22664	18,5%

- 8 reviewers
- 100 random links per strategy
- nouns, verbs, adjectives, adverbs
- single confidence score per link based on all weighted strategies

Results of the macro-alignment

	LUS	LINKED		NOT-LINKED
RBN	66.024	47.250	72%	18.774
VLIS	106.504	46.924	44%	59.580
RBN-VLIS LINKS		58.053		

Cornetto Mapping Record

- CID unique pointer to bind them all, assigned by IRION
- C_LU_ID LU id to be assigned to each LU in CDB
- C_SY_ID SYNSET id to be assigned to each synset in CDB
- C_FORM lexical form
- C_SEQ_NR sequence number in CDB
- R_LU_ID LU id currently used in RBN
- R_SEQ_NR sequence number currently used in RBN
- D_LU_ID LU id currently used in DWN (original Vlis ID)
- D_SEQ_NR sequence number currently used in DWN
- D_SY_ID synset id currently used in DWN
- Score confidence score assigned by algorithm
- Status manually confirmed
- Name editor



Creation of Cornetto LUs and Synsets

- No mapping for a LU in RBN to a synonym in DWN:
 - create unique LU in Cornetto based on RBN LU. We do not create a synset for the LU in Cornetto;
- No mapping for a synonym in DWN to an LU in RBN:
 - create unique synonym in a unique synset in Cornetto
 - create corresponding Cornetto LU with the information from DWN;
- If there is a best scoring mapping between an LU in RBN and a synonym in DWN:
 - create single unique LU and a single unique synonym in Cornetto that point to each other and to both RBN and DWN;
- All remaining mappings:
 - do not create LUs and/or synsets;
 - stored as additional mappings (as weighted alternatives);

Micro-alignment

- ❑ Separate layers with co-indexing:
 - DWN:
 - ❑ gitarist<0> -co_agent_instrument-> gitaar<1>
 - RBN:
 - ❑ gitarist <0> speelt **op** een gitaar <1>
- ❑ Unified CBN structure:
 - Event structure:
 - ❑ E: gitaarspelen <e0>
 - ❑ A1: gitarist <a1>
 - ❑ A2: gitaar <a2>
- ❑ Conceptual information shared by all synonyms
- ❑ Lexical information unique per synonym

Database design

Lexical Unit & Synsets

- Lexical Unit = form-meaning relation, such that:
 - form = abstract representation of certain realizations;
 - part-of-speech is the same;
 - meaning is the same, where meaning is defined by a reference to a unique Synset;
- Synset = Set of synonyms (LUs) that refer to the same entities in most contexts.
 - Defined by lexical semantic relations;
 - Defined by reference to ontology Terms or KIF expressions involving Terms from the ontology;

Data structure overview

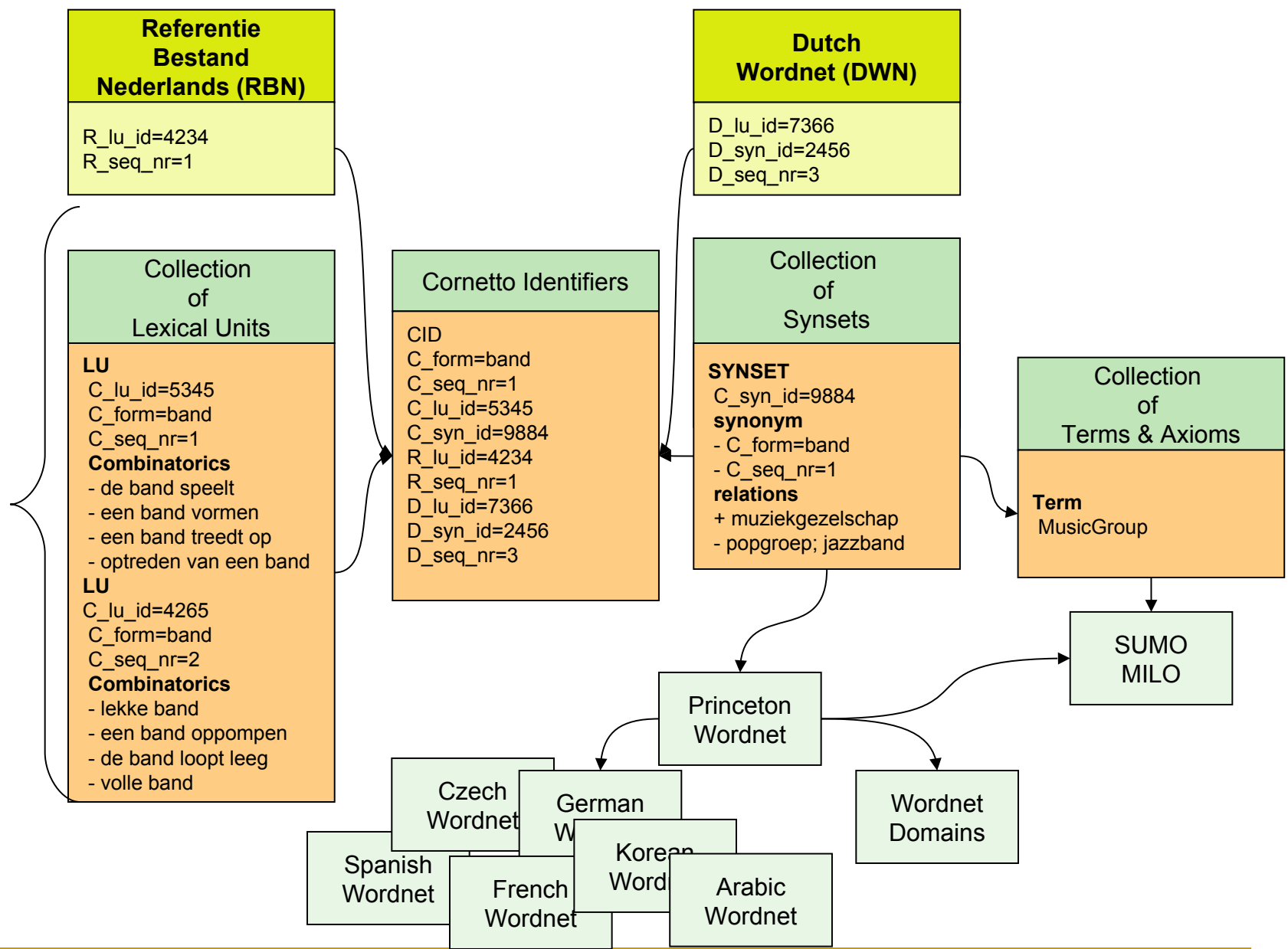
■ Collections:

- ❑ Lexical units (LU): -> mainly derived from RBN
- ❑ Synsets (SY): -> mainly derived from DWN
- ❑ Terms (TE): -> based on SUMO/MILO, linked to PWN
- ❑ Domains (DM): -> based on Wordnet domains

■ Mappings:

- ❑ LU<-> SY
- ❑ SY <-> SY (within Dutch and from Dutch to English)
- ❑ SY <-> TE
- ❑ SY <-> DM

Cornetto Database (CDB)



Disambiguating combinations

slag & leveren (fight)
slag & vrije (free style swimming)

punt & delen (share points)
punt & pakken (take a point)
punt & scoren (score a point)
punt & hard (firm point)

trap & geven (to kick)
trap & krijgen (be kicked)
trap & vrij (free kick)
trap & hoge (high stairs)

band & lekke (flat tire)
band & lopende (converter belt)
band & losse (loose relation)
band & spelen (music band plays)
band & starten (start a tape or belt)
band & sterke (strong relation)

bui & flinke (heavy shower)
bui & plaatselijk (local shower)
bui & vallen (there can be shower)
bui & slechte (bad mood)

Prototypical contexts

functie & bekleden (take a position)
functie & hoge (high position)
functie & neerleggen (put down position)
functie & ontheffen (discharge from position)
functie & openen (open up a position)
functie & treden (start a position)
functie & uitoefenen (carry out position)
functie & vervullen (fulfill position)

werk & aannemen (take on work)
werk & bieden (offer work)
werk & gaan & naar (go to work)
werk & half (incomplete work)
werk & staken (strike)
werk & zetten & aan (to put to work)
werk & zoeken (search for work)
werk & zwaar (heavy work)

Semantics for frame structures

- Event structure for verbs from RBN:
 - ❑ E: behandelen <e0> action
 - ❑ A1: <?a1> pers
 - ❑ A2: <?a2>
 - ❑ C3: <?c3> prep
 - ❑ iemand aan [zijn verwondingen] behandelen
 - ❑ een patiënt voor [een nieraandoening/puistje/keelpijn] behandelen
 - ❑ iemand met [fysiotherapie/medicijnen]_{Instrument} behandelen
- DWN:
 - ❑ [causes] [v] genezen:2, beteren:1, herstellen:1
 - ❑ [involved_agent] [n] arts:1; dokter:1 <?a1>
 - ❑ [involved_patient] [n] zieke:1; patiënt:1 <?a2>
 - ❑ [involved_instrument] [n] hart-longmachine:1 <?c3>
 - ❑ [involved_instrument] [n] mitella:1, draagdoek:1 <?c3>
 - ❑ [involved_instrument] [n] geneesmiddel:1; medicijn:1 <?c3>
 - ❑ etc...

Advantage of combining RBN & DWN

- Missing information can be derived from the other resource;
- Argument structures & frames of RBN can be enriched with semantics => WSD;
- Add hierarchy to combinatorics;
- Synsets in DWN are enriched with contextual information: => WSD;
- Synsets in DWN are enriched with lexical & grammatical selection criteria: => text generation;

Ontologize Cornetto

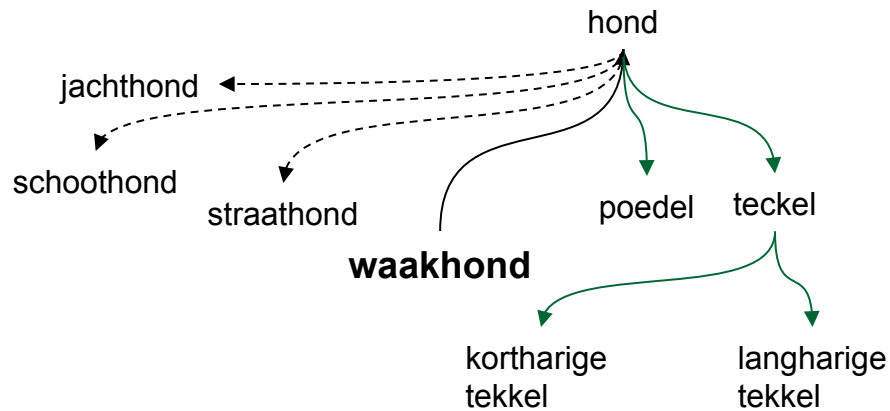
- Identity criteria OntoClean (Guarino & Welty 2002), :
 - **rigidity**: to what extent are properties true for entities in all worlds? You are always a *human*, but you can be a *student* for a short while.
 - **essence**: what properties are essential for an entity? *Shape* is essential for a *statue* but not for the *clay* it is made of.
 - **unicity**: what represents a whole and what entities are parts of these wholes? An *ocean* is a whole but the *water* it contains is not.
- Hyponyms of *hond* (dog) in DWN:
 - bokser; corgi; loboer; mopshond; pekinees; pointer; spaniël;
 - pup; reu; teef
 - bastaard; straathond; blindengeleidehond; bullebijter; diensthond; gashond; jachthond (hunting dog); lawinehond; schoothondje (lap dog); waakhond (watch dog)

Identity criteria applied to DWN

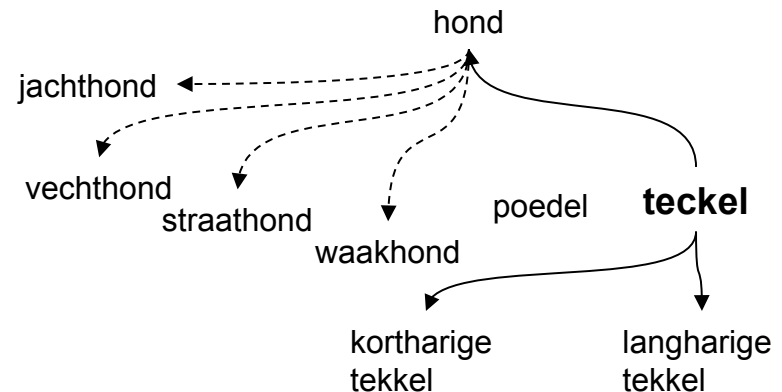
- (Semi-)rigid type hierarchy in the ontology:
 - Canine => PoodleDog; NewfoundlandDog; DalmatianDog, etc.
- Wordnet consists of names for (semi-)rigid dog-types and other words for dogs with roles:
 - poedel = PoodleDog
 - jachthond (?CAN)
 - ⇔ (and
(instance ?CAN Canine)
(instance ?EV Hunting)
(agent ?CAN ?EV)))
- Type hierarchy remains compact and pure

Expansion with pure hyponymy relations

Expansie van rol naar types

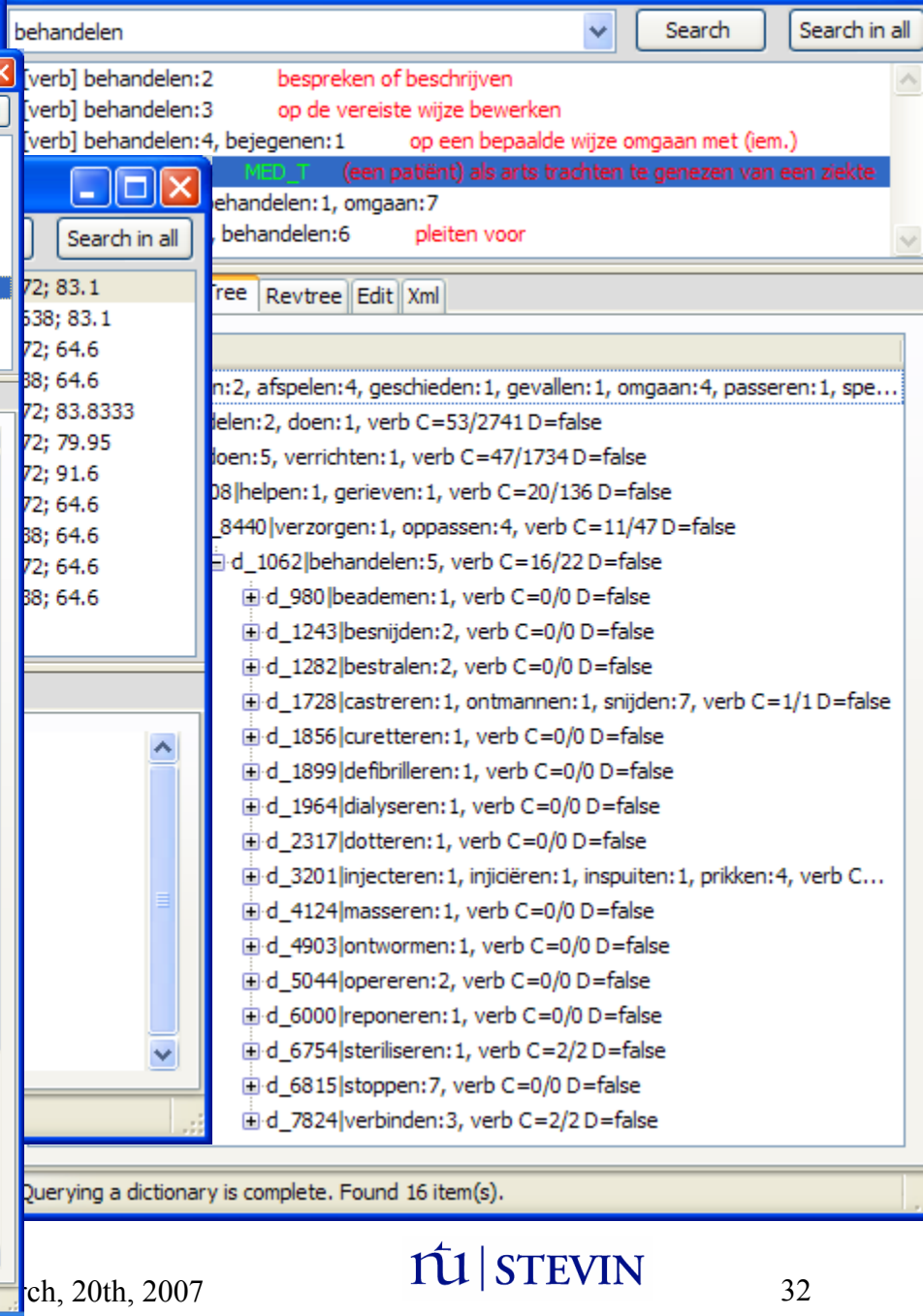


Expansie van types naar rollen



Advantage of ontologizing Cornetto?

- More explicit definition of hierarchical structures & roles: => **better expansion**;
- Apply semantic implications through relations results in more exact database;
- Enables inferencing;
- Enables the development of domain specific database: => **required for semantic web applications**;



behandelen

Search

Search in all

verb, behandelen:1
 verb, behandelen:1, omgaan met iets of iemand
 verb, behandelen:2
 verb, behandelen:2, bespreken
 verb, behandelen:3
 verb, behandelen:3, trachten te genezen
 verb, behandelen:4
 verb, behandelen:5
 verb, behandelen:6

Form Syn-Sem Examples Edit Xml

New

Delete

Save

Clear form

Lu id

r_v-1172

Seq. nr.

3

Form

Spelling behandelen

Length

Spelvar

Morphology

Type derivation

Structure

Flex mode inf

Tense

ntense

Number nnumber

Person

nperson

Flex conjugation

Add

regular

Remove

Syntax

Trans Tran

Separ

Onsch

Class Main

Peraux

H

Valency Tri

Reflexiv

Nrefl

Subject Pers

Complementation

Add

nil

Remove



behandelen



Search

Search in all

verb, behandelen:1
verb, behandelen:1, omgaan met iets of iemand
verb, behandelen:2
verb, behandelen:2, bespreken
verb, behandelen:3
verb, behandelen:3, trachten te genezen
verb, behandelen:4
verb, behandelen:5
verb, behandelen:6

Form Syn-Sem Examples Edit Xml

```
- <cdb_lu c_seq_nr="3" type="swu" c_lu_id="r_v-1172">
  <form form-cat="verb" form-spelling="behandelen"/>
  - <morphology_verb>
    <morpho-type>derivation</morpho-type>
    - <flex-conjugation>
      <flex-conjugationtype>regular</flex-conjugationtype>
    </flex-conjugation>
    <flex-mode>inf</flex-mode>
    <flex-tense>ntense</flex-tense>
    <flex-number>nnumber</flex-number>
    <flex-person>nperson</flex-person>
  </morphology_verb>
  - <syntax_verb>
    <sy-trans>tran</sy-trans>
    <sy-separ>onsch</sy-separ>
    <sy-class>main</sy-class>
    <sy-peraux>h</sy-peraux>
    <sy-valency>tri</sy-valency>
    <sy-reflexiv>nrefl</sy-reflexiv>
    <sy-subject>pers</sy-subject>
    - <sy-complementation>
      <sy-comp>np</sy-comp>
      <sy-comp>pp [...] np</sy-comp>
      <sy-compl-text>iemand voor /met iets behandelen</sy-compl-text>
    </sy-complementation>
  </syntax_verb>
  - <semantic_verb>
```

behandelen

▼

Search

Search in all

[verb] behandelen:2 bespreken of beschrijven

[verb] behandelen:3 op de vereiste wijze bewerken

[verb] behandelen:4, bejegenen:1 op een bepaalde wijze omgaan met (iem.)

[verb] behandelen:5 MED_T (een patiënt) als arts trachten te genezen van een ziekte

[verb] hanteren:1, behandelen:1, omgaan:7

[verb] verdedigen:2, behandelen:6 pleiten voor

Concept

Relations

Tree

Revtree

Edit

Xml

New

Delete

Save

Clear form

Synset Id

d_1062

☐ Base concept

Synonyms: Cid Id, Lu Id, Status

Add

168519

▼

verb, behandelen:5

▼

Remove

Domain Relations

Add

Semantic Relations: Target, Relation

Add

[noun] arts:1, dokter:1, esculaap:2, geneesheer:1, ▼

INVOLVED_AGENT

▼

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] beademen:1 | MED_T | (een patiënt) kunstmat ▼

HAS_HYPONYM

▼

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[noun] behandeling:2 | | het omgaan met iets ▼

XPOS_NEAR_SYNONYM

▼

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] besnijden:2 | MED_T | de voorhuid wegnemen ▼

HAS_HYPONYM

▼

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] bestralen:2 | CUL, IND | produkten met behulp ▼

HAS_HYPONYM

▼

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] castreren:1, ontmannen:1, snijden:7 | MED_T ▼

HAS_HYPONYM

▼

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] curetteren:1 | MED_T | (de baarmoeder) schof ▼

HAS_HYPONYM

▼

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

Querying a dictionary is complete.

behandelen		Search	Search in all
[verb] behandelen:2	bespreken of beschrijven		
[verb] behandelen:3	op de vereiste wijze bewerken		
[verb] behandelen:4, bejegenen:1	op een bepaalde wijze omgaan met (iem.)		
[verb] behandelen:5	MED.T (een patiënt) als arts trachten te genezen van een ziekte		
[verb] hanteren:1, behandelen:1, omgaan:7			
[verb] verdedigen:2, behandelen:6	pleiten voor		

Concept Relations Tree Revtree Edit Xml

```
- <cdb_synset c_sy_id="d_1062" count_childs="22" d_synset_id="d_1062">
- <synonyms>
  <synonym status="" c_cid_id="168519" form="behandelen" c_lu_id="d_24306"/>
</synonyms>
- <wn_internal_relations>
- <relation factive="" reversed="false" relation_name="INVOLVED_AGENT" negative="false" coordinative="false" disjunctive="false" target="d_10091">
  <author name="Piek" score="0.0" status="YES" date="19961217" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_980">
  <author name="Paul" score="0.0" status="YES" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="XPOS_NEAR_SYNONYM" negative="false" coordinative="false" disjunctive="false" target="d_15850">
  <author name="Piek" score="0.0" status="YES" date="19961129" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1243">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1282">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1728">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1856">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1899">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
```

Language technology opportunities

Text analysis

- Disambiguate word meanings in text:
 - horizontal relations
 - frame structurea with combinatoric constraints
 - examples
- More refined expansion to variants: more differentiated hyponymy-structures;
- Better definition of cohesion because we can recognize lexical inclusion:
 - knippen -> schaar
 - genezen -> ziekte, verwonding
- Inferencing, e.g. time-dependencies:
 - geboorte -> huwelijk -> scheiding -> sterven
 - oprichten -> muziekoptreden -> uit elkaar gaan

Text generation

- Selection of common combinaties from alternatives that are generated by wordnet or ontology:
 - drank -> prepareren; maken; brouwen
 - limonade maken; koffie zetten; drankje brouwen
 - smaakvolle soep; pittig eten; sterke koffie;
- Realisation of conceptual relations in syntactically correct constructions:
 - behandelen aan verwondingen
 - behandelen voor een ziekte

User scenarios

Applications in Cornetto

- Automatic acquisition of new concepts and relations from text corpora;
- Extension of the database (words & ontology) to the legal domain;
- Improve performance of existing systems;
 - Q&A
 - Multimedia & text retrieval
 - Dialogue systems

Q&A systeem UVA

- Gebruikt wordnet synsets voor:
 - Classificatie van vragen;
 - Parafrasering van vragen;
- Leidt expansie via disjuncte types tot beter resultaat?
- Coverage belangrijker dan diepe analyse van rijke data;
- Exploitatie van de rijkdom in Q&A system is meer onderzoeksonderwerp voor PhD project

Text retrieval

- Evaluatie raamwerk van EU-projekt MEANING (IST-2001-34460)
- Queries met ambiguïteit en parafraseringen worden gextraheerd, e.g. "police cell" (jail), "cell phone" (mobile), "nerve cell" (neuron).
- Test op verschillende indexen:
 - Index van letterlijke tekst
 - Index van tekst geëxpandeerd met wordnet
 - Index van tekst geëxpandeerd met wordnet na bepalen van betekenis
- Betekenis wordt bepaald door woorden aan domeinen te koppelen en domeinen aan wordnet (Magnini et al 2000): hoge precisie en lage recall;
- Cornetto wordt gebruikt om de recall te vergroten: frame-achtige relaties als er geen domein van toepassing is;

Future scenario's

Ontluikende toepassingen

- Data mining
 - Automatic derivation of ontologies
 - Controlled inferencing
- Co-operative dialogue systems:
 - Relation between data objects are important
 - Vagueness and ambiguity are handled
- Summarizers with better cohesion
- Text generation

 tekst

 **woord**

Popgroep Beatles uit elkaar.
de band speelde op een zomeravond

band

band#1 $\rightarrow$ muziekgezelschap

```

graph LR
    Beatles --> popgroep
    popgroep -.-> band
    band --> muziekgezelschap
    bandleden --> muzikanten
    muzikanten -.-> muziekgezelschap

```

Beatles:

- opgericht: 1962
- uit elkaar: 1969
- bandleden: John, Paul, George, Ringo



TaxEdit

File Help

Projects

Burgerloket

Burgerloket_1: 2006.11.28 AD at 08:13:25 PM CET, docs: 156, nPages=748:682 :Legal:

- 1#Aangifte overlijden.txt# nPages=3# Terms: nl:42 :Legal:0.89#Medical:0.82#;
- 2#Aangifte trouwen.txt# nPages=6# Terms: nl:54 :Legal:0.9#;
- 3#Aansprakelijkheid gemeente.txt# nPages=4# Terms: nl:33 :Legal:0.89#Finance:0.82#;
- 4#Aanwijzen monument.txt# nPages=3# Terms: nl:42 :Legal:0.84#Society:0.83#Architectuur:0.81#;
- 5#Adoptie.txt# nPages=5# Terms: nl:23 :Legal:0.88#Society:0.86#;
- 6#Adresgegevens.txt# nPages=3# Terms: nl:16 :Society:0.88#Government:0.82#Transport:0.81#;
- 7#Adresgegevens De Wegwerker.txt# nPages=10# Terms: nl:13 :Society:0.92#Transport:0.81#;
- 8#Adresgegevens gemeente Gilze.txt# nPages=2# Terms: nl:3 :Farming:0.82#Architectuur:0.81#;
- 9#Adresgegevens gemeentehuis.txt# nPages=5# Terms: nl:10 :Society:0.91#Architectuur:0.81#;
- 10#Adresgegevens milieustraat.txt# nPages=7# Terms: nl:9 :Environment:0.88#Society:0.86#;
- 11#Adresgegevens Villa Mol.txt# nPages=4# Terms: nl:7 :Government:0.88#Society:0.86#;
- 12#Afgifte rijbewijs.txt# nPages=3# Terms: nl:63 :Legal:0.88#;
- 13#Afval op riolering.txt# nPages=3# Terms: nl:23 :Environment:0.85#Legal:0.83#Society:0.86#;
- 14#Afval storten.txt# nPages=3# Terms: nl:18 :Environment:0.8#;
- 15#Afvalstoffenheffing.txt# nPages=6# Terms: nl:16 :Legal:0.83#Finance:0.82#;
- 16#Alarm installatie.txt# nPages=10# Terms: nl:22 :Legal:0.87#Technology:0.83#;
- 17#Asbestverwijdering.txt# nPages=3# Terms: nl:9 :Materials:0.81#Geology:0.81#;
- 18#Automatische incasso.txt# nPages=10# Terms: nl:17 :Finance:0.89#Legal:0.81#;
- 19#Avondwinkel.txt# nPages=3# Terms: nl:15 :Commerce:0.89#Legal:0.84#;
- 20#Belasting betalingsregeling.txt# nPages=20# Terms: nl:32 :Finance:0.88#Legal:0.82#;
- 21#Belasting kwijtschelding.txt# nPages=25# Terms: nl:70 :Finance:0.88#Legal:0.82#;
- 22#Beroepschrift.txt# nPages=5# Terms: nl:15 :Legal:0.92#;
- 23#Bewijs Nederlanderschap.txt# nPages=3# Terms: nl:31 :Legal:0.87#Government:0.82#;
- 24#Bewijs van in leven zijn.txt# nPages=3# Terms: nl:32 :Legal:0.89#Society:0.85#Finance:0.82#;
- 25#Bezwaarschrift belasting.txt# nPages=12# Terms: nl:44 :Legal:0.88#Finance:0.82#;
- 26#Bezwaarschrift tegen de gemeente.txt# nPages=3# Terms: nl:37 :Legal:0.89#Commerce:0.81#;
- 27#Bibliotheek.txt# nPages=3# Terms: nl:14 :Society:0.81#;
- 28#Bijstand voor zelfstandigen.txt# nPages=5# Terms: nl:43 :Society:0.86#Finance:0.82#;
- 29#Bijstandsuitkering.txt# nPages=4# Terms: nl:39 :Society:0.89#Finance:0.84#Legal:0.82#;
- 30#Bijzondere bijstand.txt# nPages=4# Terms: nl:20 :Legal:0.85#Society:0.85#Finance:0.82#;
- 31#Bingo organiseren.txt# nPages=3# Terms: nl:20 :Legal:0.84#Society:0.82#Finance:0.82#;
- 32#Bodemvervuiling.txt# nPages=8# Terms: nl:12 :Environment:0.86#;
- 33#Bouwvergunning.txt# nPages=4# Terms: nl:25 :Legal:0.86#Society:0.81#Architectuur:0.81#;

Burgerloket_1

Legal:0.89#Medical:0.82# Tops:173, Nodes:682, Docs:1326

Import View Edit Save Sort Clear

vergunning:0:76

- gemeente vergunning:1:4:0.8:4
- waterschap vergunning:3:2:0.8:2
- ventenvergunning:2:1:0.8:1
- jachtvergunning:6:2:0.8:2
- inritvergunning:2:3:0.8:7
 - inritvergunning geweigerd:2:2:0.8:2
 - inritvergunning verleend:2:2:0.8:2
- dergelijke vergunning:3:4:0.8:4
- vergunning vermelde:2:2:0.8:2
- terrasvergunning:28:3:0.8:3
- ventvergunning:8:2:0.8:4
 - ventvergunning geldig:6:2:0.8:2
- collectevergunning:2:2:0.8:2
- milieuvergunning:3:3:0.19:3
- bouwvergunning:3:6:0.3:8
 - plan bouwvergunning:1:2:0.8:2
- loterijvergunning:7:1:0.8:1
- vergunning afgegeven:2:2:0.8:2
- afdeling Vergunningen:2:6:0.8:6
- kapvergunning:1:2:0.8:2
- gewijzigde vergunning:2:2:0.8:2
- gebruiksvergunning:3:3:0.8:3
- horecavergunning:10:3:0.8:3
- aanlegvergunning:6:3:0.8:3
- voorwaarden vergunning:2:2:0.8:2
- Gemeentelijke vergunning:3:2:0.75:2
- nieuwe vergunning:3:2:0.38:2

Done loading thesaurus...

100%

TaxEdit

File Help **IRION**

Projects

Scope Knowledge

Scope Knowledge_1: 2006.10.31 AD at 10:25:30 AM CET, docs: 113, nPages=8298:64

1## nPages=0

2#alcnom# nPages=425# Terms: en:12694 :Chemistry:Physics;;

3#http://pubchem.ncbi.nlm.nih.gov# nPages=33# Terms: en:1102 :Government:Co

4#http://chemfinder.cambridgesoft.com/# nPages=59# Terms: en:222 :Communica

5#http://full.chemistry.uakron.edu/erd/index.html# nPages=3# Terms: en:16 :Commu

6#http://chem.sis.nlm.nih.gov/chemidplus/# nPages=23# Terms: en:236 :Chemistry

7## nPages=2595

8#http://chembank.broad.harvard.edu# nPages=1# Terms: en:5 :Education:Society;

9#http://www.citychemical.com/SearchResults.asp# nPages=7# Terms: en:18 :Com

10#http://www.novachem.kiev.ua/eng/wcp.php3?sb=C&pagenum=5&sf=# nPages=

11#http://www.dsl-chem.com/all_products# nPages=2# Terms: en:4 :Commerce:Le

12#http://www.wychem.com/catalogues/anilinesamidesandamines.htm# nPages=2

13#http://webbook.nist.gov/chemistry/# nPages=80# Terms: en:1090 :Physics:Tech

14#http://www.omegachem.com/en/Product_list.asp# nPages=33# Terms: en:113 :

15#http://www.alfa.com/alf/Product%20Indexes/Alfa_Complete/C10_idx.html# nPag

16#http://www.chembuyersguide.com# nPages=1# Terms: en:21 :Topography:Com

17#http://infochems.com/chemdb/chem_list.asp# nPages=1# Terms: en:2 :Chemis

18## nPages=0

19## nPages=0

20#ChemSearch# nPages=6# Terms: en:367 :Chemistry:Medical:Bio;;

21## nPages=783

22## nPages=192

23## nPages=0

24## nPages=607

25#http://www.made-in-china.com/products/catlist/listsubcat/114/00/mic/Chemicals

26## nPages=0

27## nPages=0

28#http://www.osha.gov/SLTC/healthguidelines/index.html# nPages=3# Terms: en:

29## nPages=28

30## nPages=76

31## nPages=61

32## nPages=0

33## nPages=10

Scope Knowledge_1

Generic# Tops:1041, Nodes:6446, Docs:13018

Import View Edit Save Sort Clear

acid::0:445

- acetoacetic acids;acetoacetic acid;acetoacetic acid shown:6:1:0.8:1
- acidic character of carboxylic acid;acidic character of carboxylic:4:2:0.8:2
- fluoroacetic acid acidic:2:2:0.8:2
- comparing acidities;comparing acid:3:3:0.8:3
- isomeric hydroxybenzoic acids:2:2:0.8:2
- stereoisomeric 2-butenedioic acids:2:2:0.8:2
- ascorbic acid;Ascorbic acid:4:2:0.8:2
- aldaric acid;Aldaric Acid;aldaric acids:6:1:0.8:1
- 2-Methylpentanoic Acid:2:1:0.8:1
- Benzoic Acid;Benzoic acid;benzoic acid;benzoic acids;benzoic acid's;BENZOIC ACID;Benzoic acid:4:
- Ginkgolic Acid:9:1:0.8:1
- aminoacids:2:1:0.8:1
- methyl decreases acidity of acid ten fold:2:2:0.8:2
- 2-ethylhexanoic acid;2-Ethylhexanoic Acid:3:2:0.8:2
- linoleic acid:2:2:0.8:2
- ortho-toluic acid:2:2:0.8:2
- Electron withdrawing substituents increase acidity of benzoic acid;Electron withdrawing substituents
- vigorous acid:2:1:0.8:1
- Hypobromous Acid;hypobromous acids;hypobromous acid:3:2:0.8:2
- N-Acetylneuraminic Acid:4:1:0.8:1
- dibasic acids:6:3:0.8:3
- malonic acid;malonic acids;Malonic acid:9:2:0.8:2
- aqueous acid;Aqueous acid:23:14:0.8:14
- aspartic acid;Aspartic Acid;aspartic acid shown:12:3:0.8:3
- aldonic acid;Aldonic Acid;aldonic acids:3:1:0.8:1
- myristic acid:2:2:0.8:2

Done loading thesaurus...

100%

Beperkingen van gestructureerde data

- Taalgebruik van grote publiek past niet op het taalgebruik van de informatie:
 - Regelgeving:
 - **Voertuigen** op de **openbare weg** moeten worden voorzien van een geldig **kentekenbewijs** en duidelijke **markeringen**;
 - Gebruikersvraag:
 - Ik heb een **old-timer** die ik haast nooit gebruik. Moet die ook een **kentekenplaat** en **lichten** hebben als ik die op **straat** parkeer?
- Eindgebruikers willen niet een complexe indeling volgen/leren/kennen maar hun eigen perspectief

Bacheloropleidingen

Bacheloropleiding ACW

Bachelorprogramma

Beroepsperspectieven

De kunst van het vergelijken

In deeltijd

Interessante links

Introductie

Kosten

Medewerkers

Meer informatie

Nieuws

Onderzoek

Stages

Studiebegeleiding

Studievereniging

Toelating en inschrijving

Bacheloropleiding Literatuur

Bachelorprogramma

Beroepsperspectieven

Contact

Frame

In deeltijd

Interessante links

Introductie

Kosten

Medewerkers

Meer informatie

Tijdschrift frame

Toelating en inschrijving

Vakkenoverzicht

Masteropleidingen

Masteropleiding ACW

Beroepsperspectieven

In deeltijd

Interessante links

Introductie

Kosten

Masterprogramma

Medewerkers

Meer informatie

Onderzoek

Stages

Studievereniging

Toelating en inschrijving

Masteropleiding Literatuur

Beroepsperspectieven

Contact

In deeltijd

Interessante links

Introductie

Kosten

Masterprogramma

Medewerkers

Meer informatie

Onderzoek

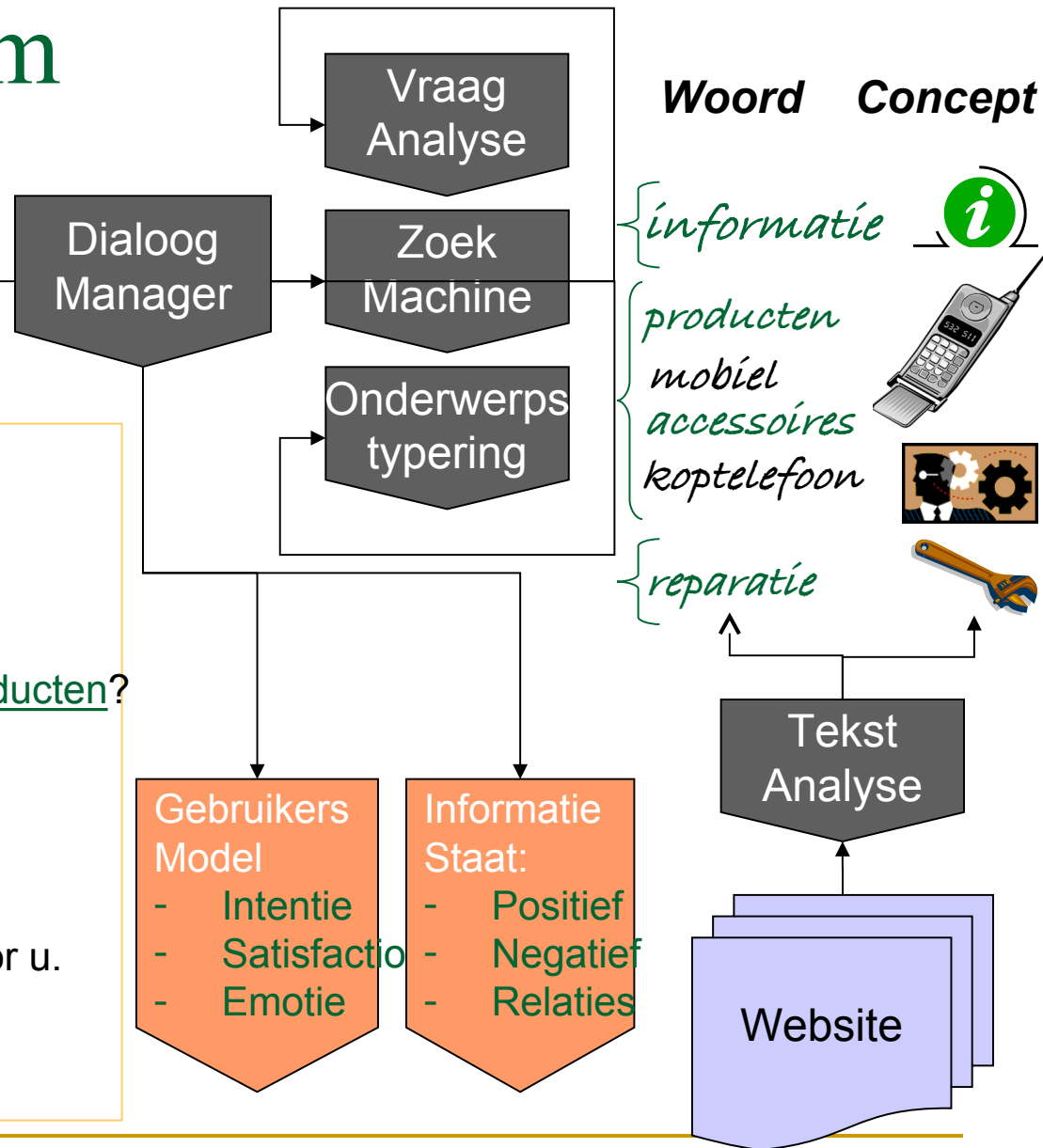
Tijdschrift frame

Toelating en inschrijving

Communicatief dialoogsysteem

- Toegang tot informatie en diensten:
 - ongeacht woordkeuze
 - ongeacht de structurering van de informatie
 - eventueel met gebruikmaking van die structurering
- Werkt samen met een gebruiker:
 - Vraagt de gebruiker om help, instructies, bevestiging, uitleg
- Gebruikt 4 informatielagen:
 - De **intentie** van de gebruiker: *klacht indienen, producten kopen, support, informatie*
 - De mate waarin iemand tevreden is (**satisfaction rate**)?
 - De **emotionele staat** van de gebruiker; *is iemand boos, vrolijk, vriendelijk?*
 - De **informatie staat** gebaseerd op de inhoudelijke beschrijving die een gebruiker geeft van een informatiebehoefte; *Waar is iemand naar opzoek?*

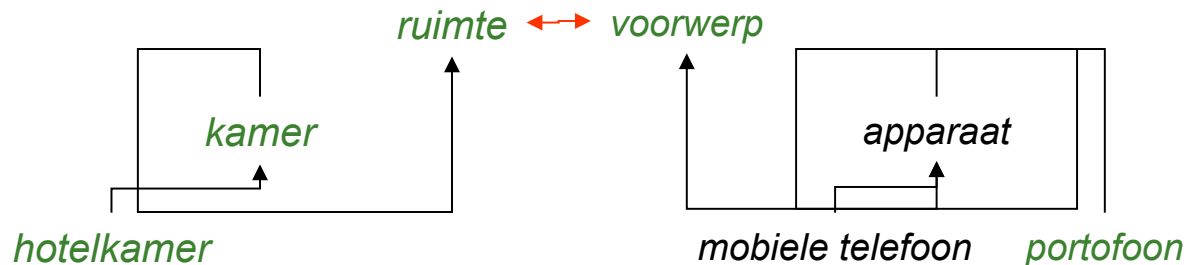
- Bekijk ze eens.



Communicatief dialoogsysteem

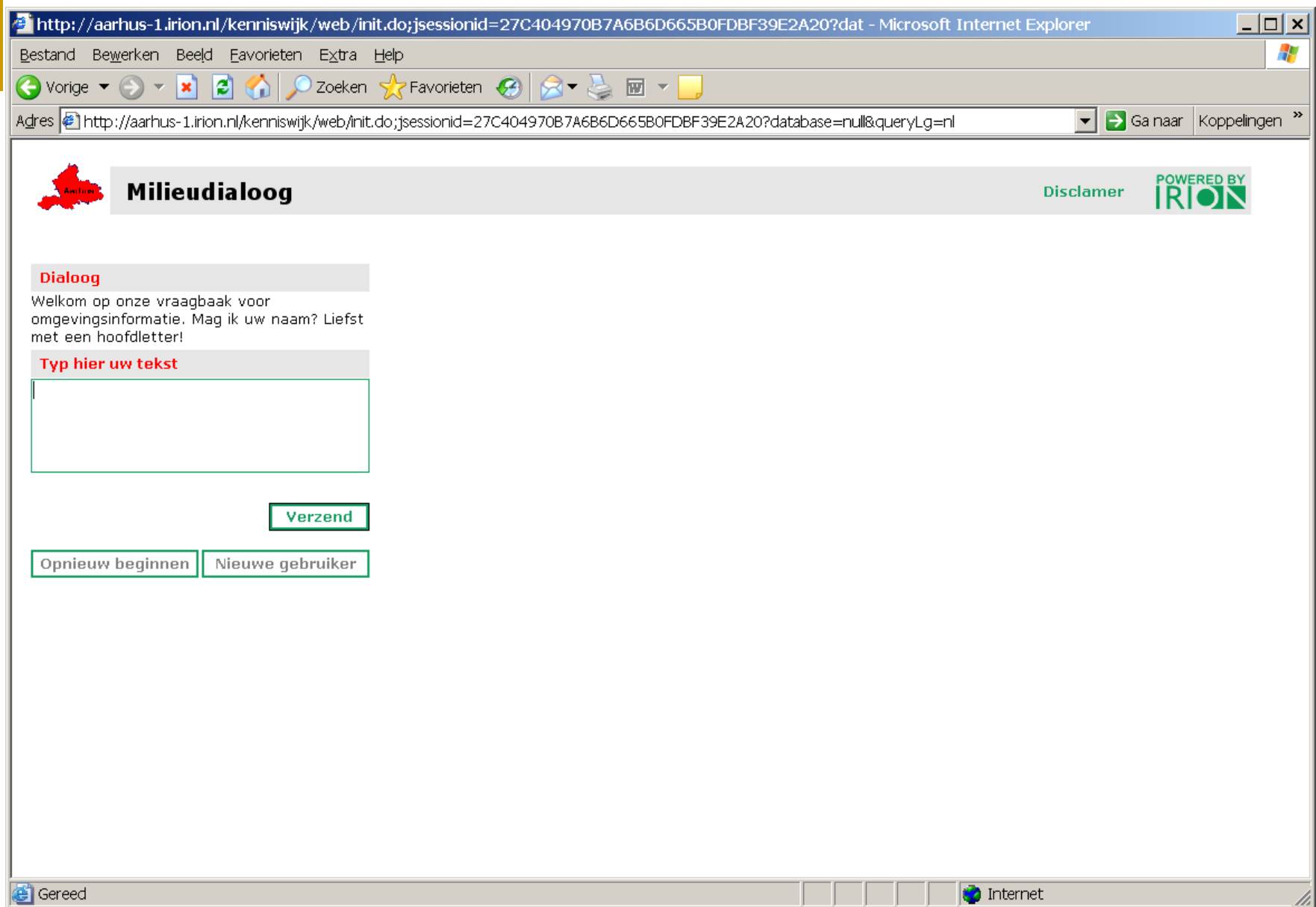
■ Voorkomt deadlocks:

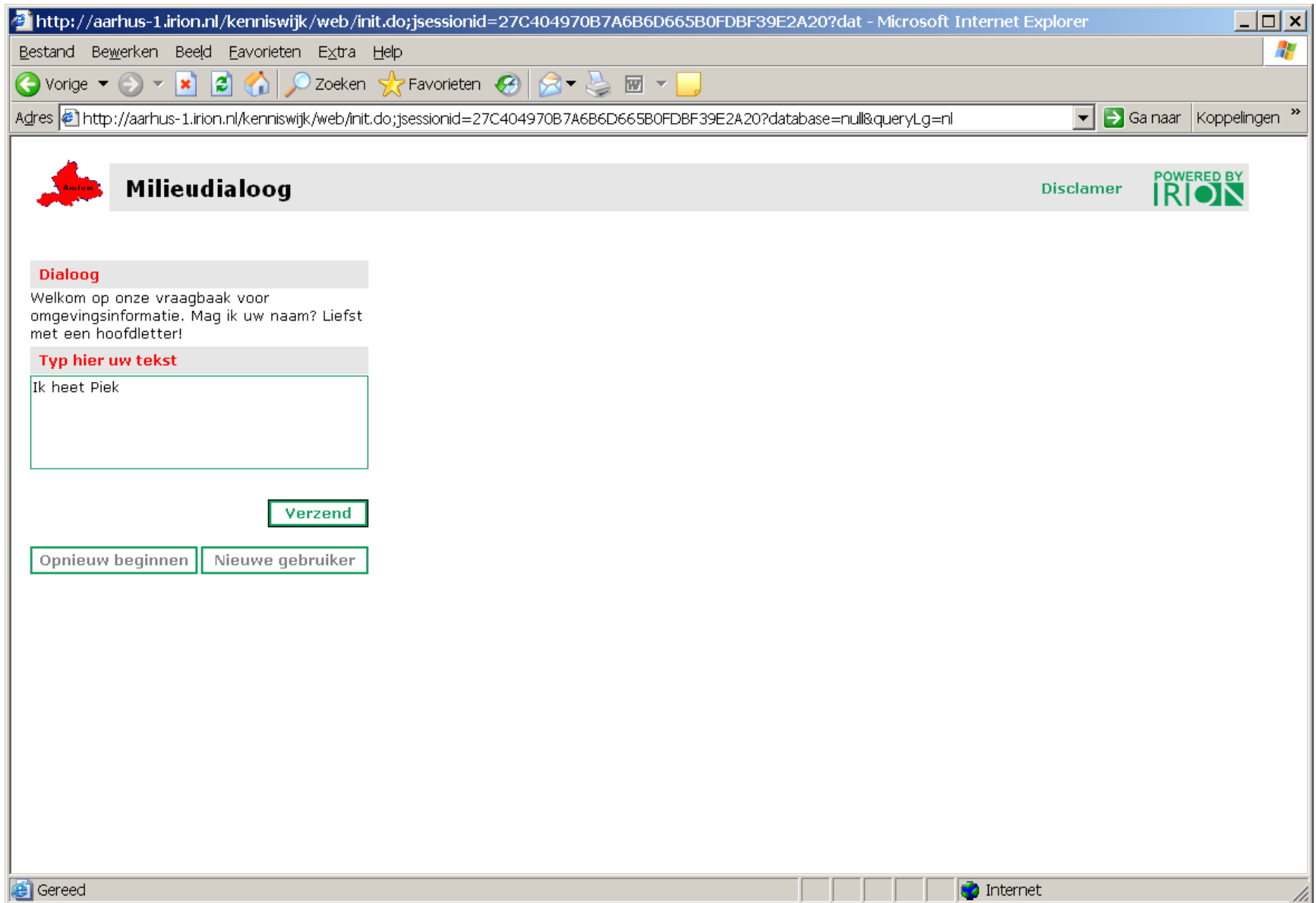
- Detecteert vaagheid en ambiguïteit (welke betekenis van *band*?)
- Detecteert veranderingen van onderwerp
- Gebruikt negatieve informatie: "*Geen muziekbond, ik zoek fietsbanden!*"
- Kan out-of-domain vragen aan:
 - "*We hebben geen hotelkamers maar wel elektronische apparaten*".
 - "*Nee, we hebben geen portofoons maar wel andere elektronische apparaten zoals mobiele telefoons*"

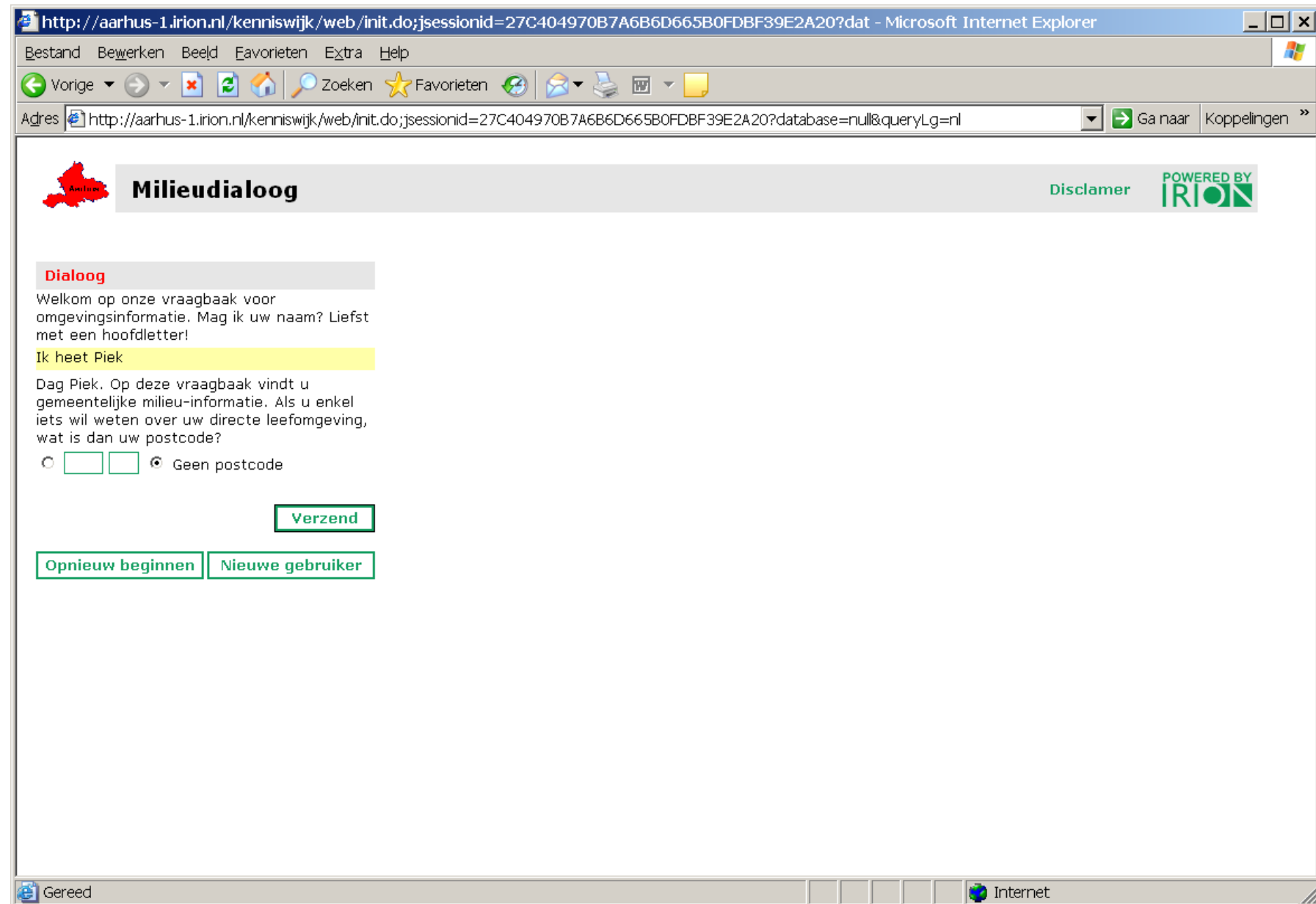




The end....







http://aarhus-1.irion.nl/kenniswijk/web/init.do;jsessionid=27C404970B7A6B6D665B0FDBF39E2A20?dat - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige Zoeken Favorieten

Adres http://aarhus-1.irion.nl/kenniswijk/web/init.do;jsessionid=27C404970B7A6B6D665B0FDBF39E2A20?database=null&queryLg=nl Ga naar Koppelingen »

 **Milieudialoog** Disclaimer 

Dialoog

Welkom op onze vraagbaak voor omgevingsinformatie. Mag ik uw naam? Liefst met een hoofdletter!

Ik heet Piek

Dag Piek. Op deze vraagbaak vindt u gemeentelijke milieu-informatie. Als u enkel iets wil weten over uw directe leefomgeving, wat is dan uw postcode?

Geen postcode.

Dank u, ik zal over de hele gemeente gaan zoeken. Ik werk als volgt: u stelt een vraag, ik leid u naar een rubriek en daarbinnen naar een document waarin u het antwoord kunt lezen. Wat wilt u weten?

Typ hier uw tekst

Verzend

Opnieuw beginnen **Nieuwe gebruiker**

Gereed Internet



http://aarhus-1.iron.nl/kenniswijk/web/init.do;jsessionid=27C404970B7A6B6D665B0FDBF39E2A20?dat - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige Zoeken Favorieten

Adres http://aarhus-1.iron.nl/kenniswijk/web/init.do;jsessionid=27C404970B7A6B6D665B0FDBF39E2A20?database=null&queryLg=nl Ga naar Koppelingen »

 **Milieudialoog** [Disclaimer](#) **POWERED BY IRON**

Dialoog

Welkom op onze vraagbaak voor omgevingsinformatie. Mag ik uw naam? Liefst met een hoofdletter!

Ik heet Piek

Dag Piek. Op deze vraagbaak vindt u gemeentelijke milieu-informatie. Als u enkel iets wil weten over uw directe leefomgeving, wat is dan uw postcode?

Geen postcode.

Dank u, ik zal over de hele gemeente gaan zoeken. Ik werk als volgt: u stelt een vraag, ik leid u naar een rubriek en daarbinnen naar een document waarin u het antwoord kunt lezen. Wat wilt u weten?

Typ hier uw tekst

Kan fijn stof tot kanker leiden?

Verzend

Opnieuw beginnen **Nieuwe gebruiker**

Gereed Internet



http://aarhus-1.irion.nl/kenniswijk/web/newSession.do - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige Zoeken Favorieten

Adres http://aarhus-1.irion.nl/kenniswijk/web/newSession.do Ga naar Koppelingen

 **Milieudialogo** Disclaimer 

omgevingsinformatie bezoekt. Wat is uw naam? Graag met een hoofdletter!

Piek

Goedendag Piek. Deze vraagbaak bevat lokale milieu-informatie. Als u enkel iets wilt vragen over uw directe leefomgeving, wat is dan uw postcode?

Geen postcode.

Oké, ik ga voor u op zoek in de hele gemeente. Ik werk als volgt: u stelt een vraag, ik leid u naar een rubriek en daarbinnen naar een document waarin u het antwoord kunt lezen. Wat wilt u vragen?

Kan fijn stof tot kanker leiden?

Ik kan helaas zelf geen vragen beantwoorden Piek. Wel kan ik wat informatie voor u opzoeken.

Wilt u iets weten over rubriek **Lucht** of **Gezondheid**?

Ga verder in Gezondheid

Oké Piek, ik zal voor u in rubriek **Gezondheid** kijken...

Ik twijfel sterk, maar misschien heeft u toch iets aan rubriek **Lucht en gezondheid**. Klik erop.

Ga verder in Lucht en gezondheid

Gelukkig Piek, ik heb iets voor u. Zie rechts.

Typ hier uw tekst

Verzend

Opnieuw beginnen **Nieuwe gebruiker**

Gevonden: 8 document(en). Klik op plusje (+) om treffers te bekijken. ?

Alles tonen **Alles verbergen**

1. CROW: Wegen naar schonere lucht (48 pag.) Metadata

88% Pag. 13 **Rubriek: Lucht en gezondheid; Postcode: 6533D;**
verbindingen tussen zuurstof en **stikstof**. De voornaamste zijn stikstofmonoxide(NO) en maximaal twee jaar eerder overlijden dan wanneer zij niet waren blootgesteld. **Fijn stof** Kan diep in de longen en verhoogde hart slag. **Mogelijk kankerverwekkend**.- Vervroegde sterfte aan alle doodsoorzaken.- **Pag in HTML**

60% Pag. 14 **Rubriek: Lucht en gezondheid; Postcode: 6533D;**
De bijdrage van de weg aan de concentraties **fijn stof** en stikstofdioxide kan wel leiden tot lokale overschrijdingen. Voor een overzicht van de werking van **fijn stof**, stikstofdioxide en ozon en blootstelling aan een hoge concentratie van **stoffen** [6]. Hierbij is onderscheid gemaakt tussen effecten bij kortdurende **Pag in HTML**

60% Pag. 38 **Rubriek: Lucht en gezondheid; Postcode: 6533D;**
bio-ethanol hebben geen effect op **fijn stof** en stikstofdioxide concentraties. Effecten op een afname van de concentraties stikstofdioxide en **fijn stof** achter het scherm. Het positieve effect neemt toe met de hoogte van het scherm. **Pag in HTML**

60% Pag. 36 **Rubriek: Lucht en gezondheid; Postcode: 6533D;**
Een bus met een aardgasmotor heeft bijvoorbeeld een duidelijk lagere uitstoot **stikstofoxiden** dan een bus met een diesel motor. Ook de uitstoot van **fijn stof** is lager. Het effect op de bijdrage van het verkeer aan **Pag in HTML**

2. Provincie Gelderland: Rapportage luchtkwaliteit 2004 (23 pag.) Metadata

3. Milieu- en Natuurplanbureau: Fijn stof nader bekeken (64 pag.) Metadata

4. Voorontwerp Uitvoeringsprogramma verbetering luchtkwaliteit Nijmegen 2005-2010 (112 pag.) Metadata

5. Rapportage Besluit Luchtkwaliteit 2002 (65 pag.) Metadata

6. Rapportage Besluit Luchtkwaliteit 2001 (65 pag.) Metadata

7. Milieu- en Natuurplanbureau: Luchtkwaliteit en Gezondheid (1 pag.) Metadata

8. Rapportage luchtkwaliteit 2004 (23 pag.) Metadata

Internet



The end....

Cornetto Synset

koffie

[noun] koffie:1
[noun] koffie:2
[noun] koffie:3, leut:2, slomp:1, slobber:3
[noun] koffieketel:1

Concept | Relations | Tree | Revtree | Edit | Xml

Synset Id ☐ Base concept

Synonyms: Cid Id, Lu Id, Status

Domain Relations

Semantic Relations: Target, Relation

Equivalence Relations: Target, Name, Version, Pos

Ontology

Querying a dictionary is complete. //

Cornetto LU

koffie

[noun] koffie:3; ,
[noun] koffie:4; ,

Form | Syn-Sem | Examples | Edit | Xml

Lu id Seq. nr. Type

Syntax

Gender Number
Article Complementation

Semantics

Definition: def, genus, specifica, defSource

Resume Def
Genus Specifica
DefSource

Reference Countability
Type Subclass
Shift Selfrestriction

Pragmatics

SocGroup Domain
Style Origin
Frequency Connotation

Querying a dictionary is complete. //

DEBVisDic English WordNet

bas

[a] bass:1, deep:6
 [n] bass:1
 [n] bass:2, bass part:1
 [n] bass:3, basso:1
 [n] bass:6, bass voice:1, basso:2
 [n] bass:7
 [n] bass:8
 [n] freshwater bass:1, bass:5
 [n] sea bass:1, bass:4

Preview Tree Revtree Edit Query

POS: n STAMP: ID: ENG
 Synonyms: bass:7

Definition: the member with
 Domain: music
 SUMO/MILO: + Musically
 -->> [hypernym] musical in
 <<-- [hyponym] bass fiddle
 contrabass:1, string bass:1
 <<-- [hyponym] bass guitar
 <<-- [hyponym] bass horn
 <<-- [hyponym] bombard

Querying a dictionary is complete.

Cornetto LU

bas Search Search in all

noun, bas:1, zangstem
 noun, bas:2, zanger
 noun, bas:3, contrabas
 noun, bas:4
 noun, bas:4, basgitaar

Form Syn-Sem Examples Edit Xml

bas

[x] [noun] bas:2,
 [x] [noun] bas:3
 [x] [noun] baspa
 [x] [noun] contra

Concept Relations

Item

d_28590|iets
 d_35105
 d_34
 d_

Querying a dictionary is complete.

C_LU_ID: r_n-6016 C_SEQ_NR: 4
 Examples:
 Form example:
 canonical form: een bas stemmen
 category: vp
 text category:
 Syntax example:
 type: fixed
 combiword:
 combicat:

Form example:
 canonical form: op bas:(Roland Knoppe)
 category: pp
 text category:
 Syntax example:
 type: fixed
 combiword:
 combicat:

Form example:
 canonical form: bas spelen

Querying a dictionary is complete.

Search in all

bijknstrument

=false

noun, bas:1, zangstem
 noun, bas:2, zanger
 noun, bas:3, contrabas
 noun, bas:4
 noun, bas:4, basgitaar

Form	Syn-Sem	Examples	Edit	Xml
New Delete Save Clear form				
Lu id r_n-6016		Seq. nr. 4		
Form Spelling bas Length shortening Spelvar				
Morphology Type simpmorph Structure Plurform Add				
Syntax Gender Mf Number Article De Complementation				
Semantics Definition: def, genus, specifica, defSource Resume basgitaar Def Genus Specifica DefSource				
Reference Common Countability count Type Artefact Subclass Shift Subrestriction				



Search

Search in

[noun] bas:2, basstem:1

[noun] bas:3

[noun] baspartij:1, bas:1 MUZ

[noun] contrabas:1, bas:4, basviool:1 MUZ het grootste en diepst gaande strijkinstrument

pt Relations Tree Revtree Edit Xml

```
<cdb_synset c_sy_id="d_33176" count_childs="0" d_synset_id="d_33176">
- <synonyms>
  <synonym status="" c_cid_id="8252" form="contrabas" c_lu_id="r_n-9963"/>
  <synonym status="" c_cid_id="73785" form="bas" c_lu_id="d_21942"/>
  <synonym status="" c_cid_id="114489" form="basviool" c_lu_id="d_22293"/>
</synonyms>
- <wn_internal_relations>
- <relation factive="" reversed="false" relation_name="HAS_HYPERONYM" negative="false" coordinative="false" disjunctive="false"
  target="d_21425">
  <author name="Laura" score="0.0" status="YES" date="19980105" source_id="d_33176"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_MERO_PART" negative="false" coordinative="false" disjunctive="false"
  target="d_27044">
  <author name="Laura" score="0.0" status="YES" date="19980105" source_id="d_33176"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPERONYM" negative="false" coordinative="false" disjunctive="false"
  target="d_31135">
  <author name="Paul" score="0.0" status="YES" date="19961206" source_id="d_33176"/>
</relation>
</wn_internal_relations>
- <wn_equivalence_relations>
- <relation relation_name="EQ_NEAR_SYNONYM" version="pwn_1_5" target20="ENG20-02705597-n" target="ENG15-02122515-n">
  <author name="Paul" score="1877.0" status="YES" date="19970903" source_id="HEURISTICS_BI"/>
</relation>
- <relation relation_name="EQ_NEAR_SYNONYM" version="pwn_1_5" target20="ENG20-02706182-n" target="ENG15-03240201-n">
  <author name="Laura" score="2058.0" status="YES" date="19980105" source_id="HEURISTICS_BI"/>
```

Cornetto LU

bas

noun, bas:1, zangstem
 noun, bas:2, zanger
 noun, bas:3, contrabas
 noun, bas:4
 noun, bas:4, basgitaar

Form Syn-Sem Examples Edit Xml

```

- <cdb_lu c_seq_nr="1" type="sv
  <form form-cat="noun" form-
- <morphology_noun>
  <morpho-type>simpmorpho
- <morpho-plurforms>
  <morpho-plurform>basse
  </morpho-plurforms>
</morphology_noun>
- <syntax_noun>
  <sy-gender>m</sy-gender>
  <sy-article>de</sy-article>
</syntax_noun>
- <semantics_noun>
  <sem-reference>common<
  <sem-countability>count</s
  <sem-type>nondynamic</se
  <sem-resume>zangstem</se
</semantics_noun>
- <pragmatics>
  <prag-domain general="tru
  </pragmatics>
</cdb_lu>
  
```

Querying a dictionary is complete.

Cornetto Synset

bas Search Search in all

[x] [noun] bas:2, basstem:1
 [x] [noun] bas:3
 [x] [noun] baspartij:1, bas:1 MUZ
 [x] [noun] contrabas:1, bas:4, basviool:1 MUZ het grootste en diepst gaande strijkins

Concept Relations Tree Revtree Edit Xml

C_SY_ID: d_33176

Internal relations:

relation name: HAS_HYPERONYM
 target: basinstrument1/d_21425
 coordinative: false, disjunctive: false, reversed: false, negative: false

relation name: HAS_HYPERONYM
 target: strijkinstrument1/d_31135
 coordinative: false, disjunctive: false, reversed: false, negative: false

relation name: HAS_MERO_PART
 target: bassnaar:1/d_27044
 coordinative: false, disjunctive: false, reversed: false, negative: false

Equivalence relations:

relation name: EQ_NEAR_SYNONYM
 version: pwn_1_5
 target: ENG20-02705597-n

relation name: EQ_NEAR_SYNONYM
 version: pwn_1_5
 target: ENG20-02706182-n

Querying a dictionary is complete.

Macro Alignment RBN - DWN

SYNSET	DEFINITION	DIFFERENTIAE	DOMAIN
baspartij_1 bas_1		die de bas zingt of speelt	#MUZ
bas_2 basstem_1	laagste mannenstem	laag, bij mannen	#MUZ
bas_3	baszanger, basspeler	met de basstem	#MUZ
contrabas_1 bas_4 basviool_1	het grootste en diepst gaande strijkinstrument	grootste en laagst klinkend	#MUZ

bas (noun) (bassen)

- 1 (count; nondynamic) [<gen-muz> zangstem] ⊢ <laagste> zangstem (BVD)
- 2 (count; human) [<gen-muz> zanger] ⊢ man met de stem van een bas (AA)
- 3 (count; artefact) [<gen-muz> contrabas] ⊢ strijkinstrument dat het grootst is en dat het laagste speelt (AA): *contrabas*
- 4 (count; artefact) [<gen-muz> basgitaar] ⊢ basgitaar (BVD-1)

6 Senses out of 8 candidates

- 1: lowest singing voice, RBN-1 & DWN-2
- 2: man with the voice of a bass, RBN-1 & DWN-3
- 3: biggest and lowest string instrument, RBN-3 & DWN-4
- 4: bass guitar, RBN-4
- 5: part of the music for the bass, DWN-1
- 6: bass singer or player, DWN-3

Macro alignment approach

- Feature match across RBN & DWN
 - Shared features DWN-RBN, DWN-PWN
 - Lemma, POS, hyperonym, definitions, domain labels, synonyms, semantic features (+/-animate)
 - Dependent features:
 - Relations, e.g. instrument <-> theme's
 - Ontology <-> syntactic complements
 - Merge tables, domain labels across resources
- Implementation of heuristics
 - Benchmarking & Normalization:
 - Samples per heuristic
 - Multiple reviewers
 - Combined probability:
 - Overall score
 - Score per heuristic
 - 60.64.56.89.67.45.34.89

Macro alignment

- Similar starting points for RBN and DWN:
 - Basis Lexicon Nederlands: 25,000 entries/lemmata.
 - INL frequency information.
 - Relevance to education and language learning.
- DWN is built following the methodology of EuroWordNet:
 - EWN Base Concepts that play an important role in wordnets of 12 European languages.
 - All important semantic concepts (needed to define more specific concepts) are included.
- Early test: 85% overlap between the vocabularies of the two resources.
- Intersection of about 35K entries, whereas the union of the two resources will represent about 50K entries.
- RBN only entries (c.a. 5K) needs to be linked to the DWN hyponymy structure
 - Match hyperonym of RBN with entry of DWN (creating a (parent) hyponym relation linked to DWN)
 - Select the appropriate sense (if necessary) by matching all the other features. This will result in an extension of the 35K to about 40K.
- DWN only entries (c.a. 10K) are part of the complete hierarchy but lack the information from RBN.
 - Relatively specific entries, mostly (compound) nouns that have a less complex argument structure and that some of the information can be acquired or derived.

Lexical Unit: subtypes

- ❑ Different data structures for nouns, verbs & adjectives
- ❑ Single word unit:
 - simple morphological structure
 - complex morphological structure
- ❑ Multiword unit (MWU):
 - idiomatic expressions (non-predictable meaning)
 - collocations
 - free combinations:
 - ❑ Non-predictable distribution
 - ❑ Non-predictable meaning specialization, e.g. “gaan koffiedrinken”
 - regular examples
- ❑ Suffixes
- ❑ Excluded are **productive, compositional** derivations, compounds and multiword expressions (possibly examples)



Lexical Unit: Meaning variants

- Roles, including male/female variants, e.g.:
 - theoloog/theologe
 - leraar/lerares
 - secretaresse vs. mannelijke secretaresse
 - kleuterleidster vs. ?

Viewpoints with respect to ROLEs

- Disjunct types = list of Terms in ontology.
 - If distinction between the types MaleTheologist and FemaleTheologist, then distinction between LUs:
 - LU “theoloog” (GENDER neutral form)
 - CO: MaleTheologist
 - LU “theologe”
 - CO: FemaleTheologist
- Ontology is minimalistic:
 - professions as abstract roles that can be combined with the disjunct Terms Man and Woman:
 - “theoloog” and “theologe” represent a single LUs from which we can derive the male and female forms with derivational rules (zero-derivation for the male variant):
 - LU “theoloog”
 - CO: Theologist
 - “theoloog” and “theologe” still represent distinct LUs where each has a complex KIF expression to differentiate the meanings:
 - LU “theoloog”
 - CO: (and (?X Man)
(?X Theologist))
 - LU: “theologe”
 - CO: (and (?X Woman)
(?X Theologist))

Cornetto Position with respect to ROLEs

- If defined exhaustively as a role that is neutral with respect to male/female AND
 - If the male/female form can be derived with a regular and compositional derivation
- THEN 1 LU for the abstract neutral form.
- In all other cases: separate LUs, possibly related to different ontology terms or a KIF expressions, depending on the ontology.

Consequences:

- Single LU for *theoloog*:
 - neutral form (possibly zero-derivation) for *man*.
 - neutral form or apply a derivational rule to create *theologe* when applied to *woman*
- Separate LUs for *leraar* (male teacher) and *lerares* (female teacher), there is no neutral form. The same applies to *verpleger* (male nurse) and *verpleegster* (female nurse).

Note: the derivational component still needs to ‘*know*’ when the female form can be generated and when not, e.g. *arts* (docter) where we do not generate a female form, others where we do (*docente*, *dichteres*, *kapper/kapster*), even when there is a neutral form.

Lexical Unit: polysemy

- Complex Types (Pustejovsky):
 - Book incorporates different meanings layers: physical object & content
 - Not a single LU but possible super-entries that group LUs that have a specific semantic relation (shifts in RBN)
- Motivation:
 - Semi-productive:
 - World-knowledge/culture-bound

walnut: refer to a wide range of objects and products that are related and derived from the walnut tree: fruit, wood, colour, taste, ice-cream, furniture, etc.

chestnut. similar object but less productive in our culture and therefore has less conventionalized meanings, even though anybody could productively create them at any time.