



Wat is Cornetto?

Cornetto (Combinatorial Relational Network voor Taal Toepassingen) is een lexicaal-semantic database voor het Nederlands. Het bevat voornamelijk algemene woorden van het Nederlands.

Cornetto omvat:

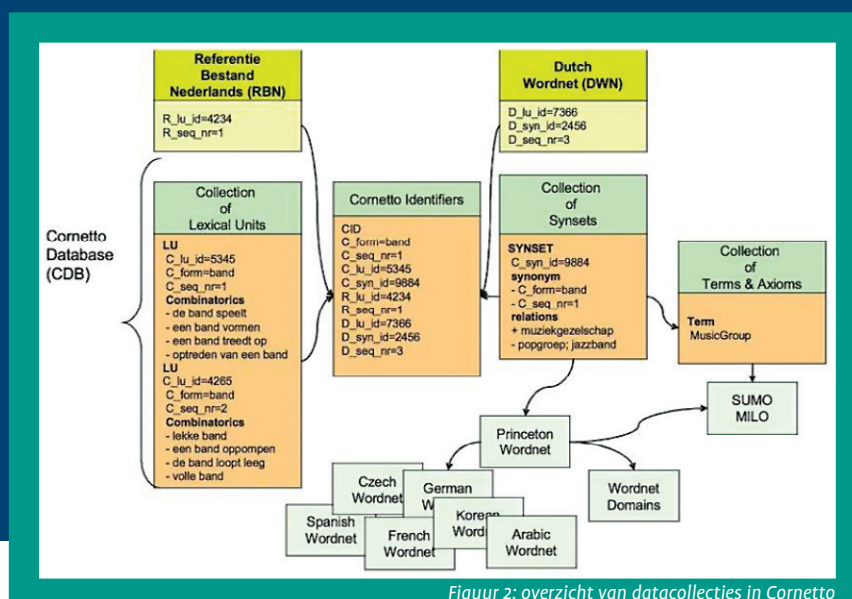
- 92.000 lemmata
- 118.000 woordbetekenissen (lexical units)
- 70.000 sets van synoniemen per woord (synsets)
- voor de meeste woorden verwijzingen naar een ontologisch raamwerk

	zelfstandige naamwoorden	werkwoorden	adjectieven	bijwoorden	andere	totaal
Synsets	52.888	9.053	7.703	220	570	70.434
Lexical units	85.278	17.363	15.731	73	21	118.466
Lemmas (form+pos)	70.556	9.055	12.307	73	n.v.t.	91.991
Synonyms in synsets	74.893	14.091	12.899	84	605	102.572
CID records (Cornetto Identifiers)	75.812	14.093	13.089	484	190	103.668
Synoniem per synset	1,42	1,56	1,67	0,38	1,06	1,46
Betekeningen per lemma	1,21	1,92	1,28	1,00	n.v.t.	1,29

Figuur 1: overzicht van de data

Bronnen

Cornetto is ontstaan uit een combinatie van twee belangrijke semantische databases voor het Nederlands: Dutch Wordnet en het Referentiebestand Nederlands. Verder is de database opgeleid met het Princeton Wordnet for English (dat weer toegang biedt tot andere wordnets), SUMO (Suggested Upper Middle Ontology) en met Wordnet Domains.



Figuur 2: overzicht van datacollecties in Cornetto



Product Sheet Cornetto

Toepassingen

Cornetto kan gebruikt worden voor verschillende computationele doeleinden, zoals tekstanalyse (bijv. desambiguering van woordbetekenissen, maken van inferenties), tekstgeneratie (bijv. selectie van gangbare combinaties uit alternatieven die worden geboden door een wordnet of de ontologie), verbetering van bestaande systemen zoals Question & Answering, multimedia, tekst retrieval, dialoogsystemen, etc.

Documentatie en software

Cornetto wordt geleverd met documentatie over het ontwerp van de database, over de relaties en alignment, etc. Er is een online demo beschikbaar waarin u de data kunt bekijken. Meer informatie vindt u op <http://www.inl.nl/nl/lexica/cornetto>.

Distributie en prijzen

Cornetto wordt door de TST-Centrale gedistribueerd als download (xml-bestanden).

prijs commercieel* op aanvraag

prijs niet-commercieel* € 0,-

**de prijzen zijn exclusief btw en eventuele verzend- en handlingkosten*

Cornetto is gefinancierd door de Nederlandse Taalunie in het kader van het STEVIN-programma. Het is tot stand gekomen door een samenwerking tussen de Faculteit der Letteren van de Vrije Universiteit Amsterdam, het Instituut voor Informatica van de Universiteit van Amsterdam, het Interdisciplinair Centrum voor Recht en Informatica van de Katholieke Universiteit Leuven, en Irion Technologies BV Delft.

Colofon

Deze Product Sheet is een uitgave van de Centrale voor Taal- en Spraaktechnologie (TST-Centrale), een initiatief van de Nederlandse Taalunie (NTU) ondergebracht bij het Instituut voor Nederlandse Lexicologie (INL). De TST-Centrale wordt gefinancierd door de NTU.

E-mail: servicedesk@inl.nl

Ontwerp: Swantje Haage Ontwerp, Amsterdam