



GÖTEBORGS
UNIVERSITET

Språk
BANKEN

CLT

Saldo-Lexin Word Sense Alignment or something like that

Elena Volodina, University of Gothenburg, Språkbanken





Resources in a nutshell

- **Definition-based** lexicon
- 29,096 senses
- 24,731 lemgrams
- Aimed at second language learners

SALDO

- **Association-based** lexicon with senses described via hierarchically-ordered associative descriptors
- 126,236 entries
- Aimed at NLP use

LEXIN ▾ 22 HITS (DISPLAYING 2)

BASEFORM	DEFINITION	PHONETIC FORM	PART OF SPEECH	RANK	LEXIN-ID	BÖJNING
data swe ↕	dator, datamaskin swe ↕	² dɑ:ta	nn	20662	3437	data ...
uppgifterna finns på data data central ↕ data spel ↕ data åldern ↕						
data swe ↕	fakta, uppgifter swe ↕	² dɑ:ta	nn	4160	3436	
skattemyndigheten begär in fullständiga data om folks inkomster ↕ data bank ↕ data bas ↕ data system data sökning						

SALDO ▾ 23 HITS

SENSE	LEMGRAM	PART OF SPEECH	PRIMARY	SECONDARY	CHILDREN (PRIMARY)
data	data (noun) 👁 237605	noun	faktum	akutvårdsdata backup dataintensiv datamaterial hålkort more...	
data ²	data ² (noun) 👁 118043	noun	dator		

Experimental workbench

Search for search item (can be empl) 2 av 21 värden valda 22 av 22 ordklasser valda

Experiment & visualize Crowdsourcing & voting

« data, noun »»

Lexin results: 2

lexin--data..1

definition: fakta, uppgifter;
English translation: data;
English synonym: facts, information

examples: skattemyndigheten begär in fullständiga data om folks inkomster

data|bank
data|bas
data|system
data|sökning
data|tjänst

theme:

antonyms:

BOWL: bank..nn, bas..nn, begära..vb, data..nn, databank..nn, **databas..nn**, datasystem..nn, datorsystem..nn, fakta..nn, **faktum..nn**, folk..nn, fullständig..av, in..ab, inkomst..nn, **om..pp**, skattemyndighet..nn, system..nn, sökning..nn, tjänst..nn, uppgift..nn

Saldo results: 2

data..1

ancestors: faktum..1; faktisk..1; vara..1;

prims: faktum..1

secs:

siblings: show siblings (20)

suggested Lexin definition from lexin--data..1 added by: *xvoloe*

fakta, uppgifter;

English translation: data;
English synonym: facts, information

BOWL: akutvård..nn, akutvårdsdata..nn, backup..nn, bearbeta..vb, behandling..nn, **databas..nn**, databearbetning..nn, databehandling..nn, datainsamling..nn, dataintensiv..av, datakommunikation..nn, datalagring..nn, datamaterial..nn, datamängd..nn, dataregister..nn, dator..nn, exakt..av, faktisk..av, **faktum..nn**, förråd..nn, hålkort..nn, hålremsa..nn, hårddata..nn, insamling..nn, jämföra..vb, kommunikation..nn, lagra..vb, lagring..nn, maskin..nn, metadata..nn, mängd..nn, mätdata..nn, mätning..nn, nålkort..nn, **om..pp**, register..nn, samköra..vb, skit..sxc, sortera..vb, vara..vb

score & BOWL overlap:

total score = **2.25**

ratio (score/min length of bowls) = **0.11**

lexical overlap (nn, nnm, nna, vb, vbm, ab, abm, av, avm), 1 point each = **2**

functional overlap (all other pos), 0.25 points each = **0.25**

semi overlap (parts in saldo mwe) 0.1 points each = **0**

[link senses](#)

Experimental measures

List of hits (1030)

remiss.nn

PrimChildren

akutvårdsdata..1 [akutvård..1]
backup..1
dataintensiv..1 [mycket..1]
datamaterial..1
hålkort..1 [lagra..1]
hålremsa..1 [lagra..1]
hårddata..1 [exakt..1]
mätdata..1 [mätning..1]

SecChildren

databas..1 [förråd..1]
databearbetning..1 [bearbeta..1]
databehandling..1 [behandling..1]

ömdome.nn

himmel.nn

stråle.nn

ljus.nn

nåtfiske.nn

vattendrag.nn

skit.nn

räv.nn

gnista.nn

omläggning.nn

premie.nn

*For details and a demo,
visit the poster*

Thank you!

SALDO



LEXIN
SPRÅKLEXIKON FÖR INVANDRARE