

Open Multilingual WordNet & NTU Multilingual Corpus (NTUMC)

Luís Morgado da Costa, Francis Bond, Le Tuan Anh, et al.

WN Tools Overview:

- OMW structure
- New Features & Linked Resources
- Editing Interfaces

NTUMC Tools Overview:

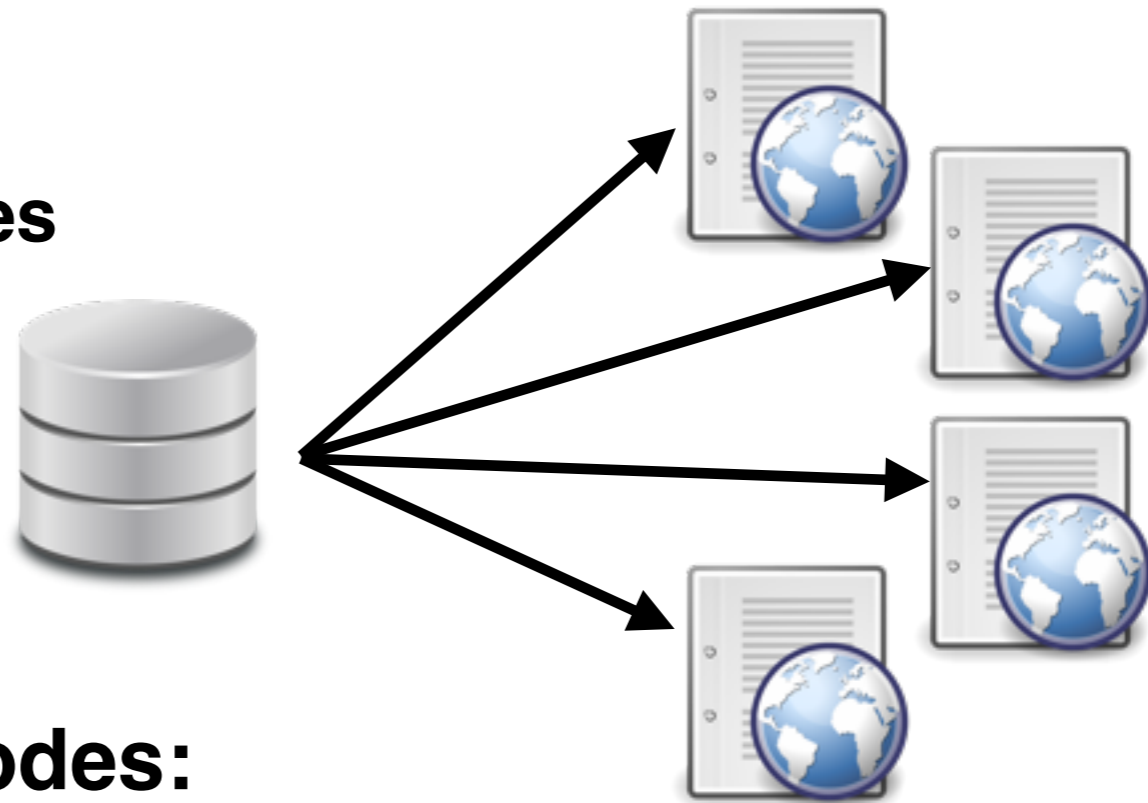
- NTUMC Structure
- Tagging Interfaces
- Support Interfaces



NANYANG
TECHNOLOGICAL
UNIVERSITY

Structure: Grid/Display Modes

**1 Database, 1 cgi :
Multiple Display Modes**



OMW Grid/Display Modes:

- + OMW (27 Languages - Human Projects)
- + OMW Extended (150+ Languages)
- + Japanese Wordnet (Japanese + English)
- + Chinese Open Wordnet (Mandarin + English)
- + WN Bahasa (Indonesian + Malay + English)
- + WN-NTUMC (Edit Mode)

New Back-Off Language Search

Results for «anjing» (ind)

02083346-n anjing	NO DEFINITION
10539715-n git, gelandangan, penipu, anjing , cacat	NO DEFINITION
02084071-n anjing	NO DEFINITION

Search WN Langs:

Results for «anjing» (ind)

02083346-n	anjing canine, canid	any of various fissiped mammals with nonretractile claws and typically long muzzles
10539715-n (4)	git, gelandangan, penipu, anjing , cacat bum, skunk, puke, stinkpot, rat, stinker, crumb, rotter, dirty dog, lowlife, scum bag, so-and-so, git	a person who is deemed to be despicable or contemptible
02084071-n (42)	anjing dog, domestic dog, Canis familiaris	a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds

Search WN Langs:

Extending & Linking Resources



150+ Languages (WIKT & CLDR) e.g. 15210045-n (January) in 143 languages

PWN Resources:

- + Core Synsets ★
- + Derivational Links
- + Lemma Frequencies
- + Verb Frames
- + Semantic Fields

External Links:

- + TempoWN
- + SentiWN
- + MLSentiCon

Synset 00829107-v ★ 'impart skills or knowledge to'

Chinese (simplified)	教导, 教授, 教训, 教育, 教
Danish	<i>lære</i>
Greek	<i>διδάσκω</i>
English	<i>teach</i> ₈₆ (▶▶▶▶), <i>instruct</i> ₉ (▶▶▶), <i>learn</i> ₂
Basque	<i>irakatsi, hezi, eskolatu</i>

Verb Frames **01864230-v** 'begin or set in motion'; V1, V2;

Something ----s; Somebody ----s something; Somebody ----s;

External Links **00020759-r** 'not ever'

TempoWN: ◀◻▶ (Past: 0.002; Present: 0.000; Future: 0.998)

SentiWN: ◻▼ (+0.00 -0.62) **MLSentiCon:** ◻▼ (+0.00 -0.25)

NTUMC-WN (Editing Mode)



**Same structure, but private:
for editing & tagging purposes**

(Currently accessible only from within
NTU, but that can change)

- + **Incremental backups** of every change made to the db;
- + The .db has been **fed with low confidence data** (that can be corrected/collected by annotators)
- + Access to parts of the **interface that allow editing** of db;
- + Requires a **valid username** and it is **timestamped**;

Edit Mode: ON

+NE

+NewSS

Multidict

1_n snow₁₃, snowfall₁
(14)

precipitation falling from clouds in the form of ice crystals

2_n snow₁₁
(11)

a layer of snowflakes (white crystals of frozen water) covering the ground

3_n Snow, C. P. Snow, Charles Percy Snow, Baron Snow of Leicester

English writer of novels about moral dilemmas in academe (1905-1980)

4_n blow, coke, nose candy, snow, C

street names for cocaine

5_v snow₁₀
(10)
v0

fall as snow

It was snowing all night

6_v bamboozle, snow, hoodwink, pull the wool over someone's eyes, lead by the nose, play false

conceal one's true motives from especially by elaborately feigning good intentions so as to gain an end

He bamboozled his professors into thinking that he knew the subject well

+NE

+NewSS

Multidict

Search WN

snow

Langs:

English

English

Current user: **lmorgado** [Change](#)

[More detail about the NTUMC+ Open Multilingual](#)

(0.01272 seconds)

Creating New Entries

Creating a new synset linked to 05650820-n

Indonesian lemmas: *bahasa pertuturan*

Chinese (simplified) lemmas: 语言

Japanese lemmas: 言語能力

English lemmas: *language*¹ *speech*

Definition: the mental faculty or power of vocal communication; the mental faculty or power of vocal communication;

New Synset Information:

You need to assign it an English name (e.g. high frequency synset), a POS (part of speech), and one or more lemmas of any language.

synset name POS

Hypernym: +

Chinese (simplified) +

Chinese (simplified) +

Lookup Word (or Synset): Language:

New Named Entity Information:

You need need to choose the NE type and to assign it a lemma.

Organization

English +

Lang:

English

Editing Existing Entries

Synset [05650820-n](#)

Add new information to this entry:

Chinese (simplified)	<input type="text" value="lemma"/>	+
English	<input type="text" value="definition"/>	+
English	<input type="text" value="example"/>	+
Change Me!	<input type="text" value="linked synset"/>	+
<input type="button" value="Add to synset"/>		

Lemmas

Indonesian	<input type="text" value="bahasa"/>	<input type="text" value="1.00"/>	<input type="button" value="✖"/>	<input type="button" value="Update"/>
	<input type="text" value="pertuturan"/>	<input type="text" value="1.00"/>	<input type="button" value="✖"/>	
Chinese (simplified)	<input type="text" value="语言"/>	<input type="text" value="0.91"/>	<input type="button" value="✖"/>	<input type="button" value="Update"/>
	<input type="text" value="言語能力"/>	<input type="text" value="1.00"/>	<input type="button" value="✖"/>	<input type="button" value="Update"/>
English	<input type="text" value="language"/>	<input type="text" value="1.00"/>	<input type="button" value="✖"/>	<input type="button" value="Update"/>
	<input type="text" value="speech"/>	<input type="text" value="1.00"/>	<input type="button" value="✖"/>	

Definitions

Japanese

<input type="text" value="口頭で意思伝達する、心的能力あるいは機能"/>	<input type="button" value="✖"/>	<input type="button" value="Update"/>
---------------------------------------------------	----------------------------------	---------------------------------------

English

<input type="text" value="the mental faculty or power of vocal communication"/>	<input type="button" value="✖"/>	<input type="button" value="Update"/>
---------------------------------------------------------------------------------	----------------------------------	---------------------------------------

Examples

Japanese

<input type="text" value="人間は言語によってほかの動物と区別される"/>	<input type="button" value="✖"/>	<input type="button" value="Update"/>
---------------------------------------------------	----------------------------------	---------------------------------------

English

<input type="text" value="language sets homo sapiens apart from all other animals"/>	<input type="button" value="✖"/>	<input type="button" value="Update"/>
--------------------------------------------------------------------------------------	----------------------------------	---------------------------------------

Relations

Hypernym:	<input type="text" value="05650329-n"/>	<input type="button" value="✖"/>
In Domain-Category:	<input type="text" value="00116203-v"/>	<input type="button" value="✖"/>
Meronym-Part:	<input type="text" value="05651068-n"/>	<input type="button" value="✖"/>
Meronym-Part:	<input type="text" value="05651242-n"/>	<input type="button" value="✖"/>
<input type="button" value="Update"/>		

Ongoing & Future WN Extensions



On the pipeline for OMW:

- + **Revise the DB Structure** to accommodate different lemma level representation (orthographic variations, use of multiple scripts, romanisation, etc.);
- + **Online Reports/Stats & Adjudication Interface** (new data);
- + **New Display Modes** (e.g. **experimental data**);
- + **Allow editing of sense links**;
- + Improve **Login System**;
- + **'Beautifying'** the interfaces (e.g. language selection, edit modes, etc.);
- + We're always open to **suggestions!**

NTUMC Corpus Tools

(It is still work in progress!)

NTUMC Tools Overview:

- NTUMC Structure
- Tagging Interfaces
- Support Interfaces



NTUMC Structure



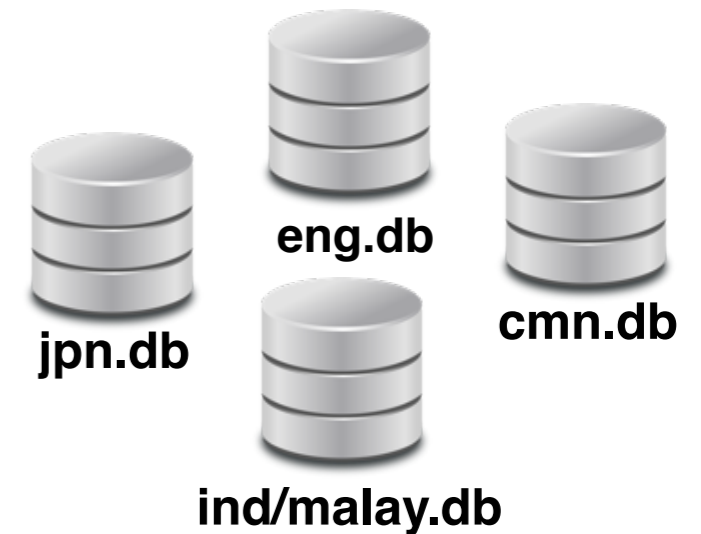
NTUMC Quick Facts:

- + Multilingual (Eng, Cmn, Jpn, Ind/Malay);
- + Multi-genre (Short Stories, Essay, News, Tourism Blog);
- + 6,000 sentences (in cmn, eng, jpn); + 2,000 (in Indonesian);
- + Hand corrected: Tokenisation, POS tagging;
- + Hand-tagged: MWE; Sense annotation; cross-lingual sense alignments;
- + Goal: Integrate with Full Deep Semantics (by HPSG Grammars);

NTUMC Structure

One database per language (3 main levels):

- + Sentence Surface Form
- + Word (Token, POS, Lemma)
- + Concepts (Word Links, Lemma, Tag)



(SID:11025) "How absurdly simple!" I cried.

(SID:11026) "Quite so!" said he, a little nettled.

(SID:11027) "Every problem becomes very childish when once it is explained to you."

wid:0	wid:1	wid:2	wid:3	wid:4	wid:5	wid:6	wid:7	wid:8	wid:9	wid:10	wid:11	wid:12	wid:13
"	Every	problem	becomes	very	childish	when	once	it	is	explained	to	you	.
(Fe)	(DT)	(NN)	(VBZ)	(RB)	(JJ)	(WRB)	(RB)	(PRP)	(VBZ)	(VBN)	(TO)	(PRP)	(Fp)
["]	[every]	[problem]	[become]	[very]	[childish]	[when]	[once]	[it]	[be]	[explain]	[to]	[you]	[.]
cid:0	cid:1	cid:2	cid:3	cid:4	cid:5	cid:6	cid:7	cid:8	cid:9	cid:10			
[1]	[2]	[3]	[4]	[5]	[7]	[9]	[10]	[6]	[8]	[12]			
every	problem	become	very	childish	once	be	explain	when	it	you			
02269794-a	14410605-n	02626604-v	00031899-r	01493016-a	00181342-r	02604760-v	00939277-v	None	None	None			

(SID:11028) Here is an unexplained one.

(SID:11029) See what you can make of that, friend Watson."

Tag Interface 1 (by word)



«狗» in cmn (cmn): Search All Multidict ? Current user: lmorgado

Logout

Default: 1_ 2_ e x w Org Loc Per Dat Oth Num Year

Distribution (8): [1_ \(87%\)](#) - [2_ \(12%\)](#)

dog₄₂ **domestic dog₀** **Canis familiaris₀** ([02084071-n](#)) a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds *the dog barked all night*

[61616](#)

据认为，现在的搜查重点遗弃尸体事件，是嫌疑犯销售商向公司董事提出“繁殖狗_n可以赚钱”，于是公司董事决定购买一对非洲原产大型犬，但是只弄到手一条，故要求解除合同和退还部分费用，此时发生了纠纷。

1_ 2_ e x w Org Loc Per Dat Oth Num Year

[61618](#)

饲养狗_n等宠物可以丰富心灵。

1_ 2_ e x w Org Loc Per Dat Oth Num Year

([饲养狗_x](#))

[61622](#)

据说，近几年在发行狗血统证书的日本养狗协会登记的纯种狗_n数量增加，前年超过100个品种，达到约34万条，增加到6年前的1.5倍。

1_ 2_ e x w Org Loc Per Dat Oth Num Year

[61622](#)

据说，近几年在发行狗_n血统证书的日本养狗协会登记的纯种狗数量增加，前年超过100个品种，达到约34万条，增加到6年前的1.5倍。

1_ 2_ e x w Org Loc Per Dat Oth Num Year

[61624](#)

这种外国产的狗也有投机性，因此从事繁殖这种狗_n的人也增加了。

1_ 2_ e x w Org Loc Per Dat Oth Num Year

[61624](#)

这种外国产的狗_n也有投机性，因此从事繁殖这种狗的人也增加了。

1_ 2_ e x w Org Loc Per Dat Oth Num Year

[61626](#)

由于狗_n很可爱，再加上可以发财的话，即使有人动了贪心也不能一概指责。

1_ 2_ e x w Org Loc Per Dat Oth Num Year

pooch₀ **doggie₀** **doggy₀** **barker₀** **bow-wow₀** ([02084732-n](#)) informal terms for dogs

[131](#)

这样我们就有足够的时间像只狗_n一样周旋其间。

1_ 2_ e x w Org Loc Per Dat Oth Num Year

tag

Results for «狗» (cmn)

[02084071-n](#) (42) **犬, 狗**
dog, domestic
dog, Canis
familiaris
a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds



[02084732-n](#)
小狗, 狗, 汪汪
pooch,
doggie,
doggy, barker,
bow-wow
informal terms for dogs



Search WN Langs:

Current user: [lmorgado](#) [Change](#) [More detail about the](#) (0.01138 seconds)

[NTUMC+ Open Multilingual Wordnet \(0.9\)](#)

This project is now integrated in the [Extended Open Multilingual Wordnet \(0.9\)](#)

Maintainer: [Francis Bond](#) <bond@ieee.org>

Tag Interface 2 (by sentence)



Current user: [lmorgado](#)

Tagging country (60330:23 eng)

60327 We have put together a special feature , including introductions of the recipients ' work by members of our advisory and nominating committees , their curriculum vitae , and congratulatory messages .

60328 Times Square , echoing with construction noise , is covered in snow .

60329 This is Richmond Hill , one of a cluster of satellite cities in north Toronto , Canada , that make up the greater Toronto area .

60330 At the end of last year , the population stood at 87,000 people , but with the highest population growth rate in the country , it has now surpassed 100,000 .

60331 Times Square , located along the main avenue in town , Highway 7 , is a community-oriented shopping center whose tenants are all " Hong Kong merchants " .

60332 Yet another ' mini ' Hong Kong is attempting to be born .

60333 With the 1997 reversion of Hong Kong to China on the visible horizon and troubled by the Tianamen Square Massacre , the peak year of Hong Kong emigrants heading for Canada was 1992 .

country 1_n 2_n 3_n 4_n 5_n e x w Org Loc Per Dat Oth

Num Year

Goto sid: Lookup word:

[Documentation](#)

Multidict

1_n country₆₈, nation₃₃,
(125) state₂₁,
commonwealth₂,
land₁, res publica,
body politic

a politically organized
body of people under a
single government

*the state has elected a new
president; African nations;
students who had come to
the nation's capitol; the
country's largest
manufacturer; an
industrialized land*

2_n country₂₉, land₇, state
(36)

the territory occupied by a
nation

*he returned to the land of
his birth; he visited several
European countries*

3_n country₁₂, nation₁₀,
(24) land₂

the people who live in a
nation or country

*a statement that sums up
the nation's mood; the news
was announced to the
nation; the whole country
worshipped him*

4_n country₁₁, rural area₃
(14)

an area outside of cities
and towns

*his poetry celebrated the
slower pace of life in the
country*

5_n area₁₄₆, country₃
(149)

a particular geographical
region of indefinite
boundary (usually serving
some special purpose or
distinguished by its
people or culture or
geography)

*it was a mountainous area;
Bible country*

Multidict

Search WN

Langs:

[More detail about the NTUMC+ Open Multilingual](#)

(0.01187 seconds)

[Wordnet \(0.9\)](#)

This project is now integrated in the [Extended Open Multilingual](#)

[Wordnet \(0.9\)](#)

Maintainer: [Francis Bond](#) <bond@ieee.org>

Corpus Editing (semi-functional)

Edit Sentence: 11027

Edit sentence:

"Every problem becomes very childish
when once it is explained to you.

None

Edit words:

wid:0 ✖ [] Fe ? None	wid:1 ✖ Every DT ? every None	wid:2 ✖ problem NN ? problem None	wid:3 ✖ becomes VBZ ? become None	wid:4 ✖ very RB ? very None	wid:5 ✖ childish JJ ? childish None	wid:6 ✖ when WRB ? when None	wid:7 ✖ once RB ? once None	wid:8 ✖ it PRP ? it None	wid:9 ✖ is VBZ ? be None	wid:10 ✖ explained VBN ? explain None	wid:11 ✖ to TO ? to None	wid:12 ✖ you PRP ? you None	wid:13 ✖ . Fp ? . None
---------------------------------------	--------------------------------------------------	------------------------------------------------------	------------------------------------------------------	------------------------------------------------	--------------------------------------------------------	-------------------------------------------------	------------------------------------------------	---------------------------------------------	---------------------------------------------	----------------------------------------------------------	---------------------------------------------	------------------------------------------------	-------------------------------------------

Edit concepts:

cid:0 ✖ [1] every ? 02269794-a ? None; oldcid=300	cid:1 ✖ [2] problem ? 14410605-n ? None; oldcid=301	cid:2 ✖ [3] become ? 02626604-v ? None; oldcid=302	cid:3 ✖ [4] very ? 00031899-r ? None; oldcid=303	cid:4 ✖ [5] childish ? 01493016-a ? None; oldcid=304	cid:5 ✖ [7] once ? 00181342-r ? None; oldcid=305	cid:6 ✖ [9] be ? 02604760-v ? None; oldcid=306	cid:7 ✖ [10] explain ? 00939277-v ? None; oldcid=307	cid:8 ✖ [6] when ? None ? None	cid:9 ✖ [8] it ? None ? None	cid:10 ✖ [12] you ? None ? None
-------------------------------------------------------------------------	---------------------------------------------------------------------------	--------------------------------------------------------------------------	------------------------------------------------------------------------	----------------------------------------------------------------------------	------------------------------------------------------------------------	----------------------------------------------------------------------	----------------------------------------------------------------------------	---------------------------------------------------	-------------------------------------------------	----------------------------------------------------

+Word Save Imorgado

Online Corpus Edition (Ongoing)



On the pipeline for NTUMC:

- + **Crosslingual Tagging Interface** (semi-functional)
- + **Display Modes Annotated Corpora** (search by sense/word)
- + **Online Management/Addition of New Texts to the db;**
- + **'Beautifying'** the interfaces;
- + We're always open to **suggestions!**

Open to discussion!

Thank You

Luís Morgado da Costa
luis.passos.morgado@gmail.com



NANYANG
TECHNOLOGICAL
UNIVERSITY