CWN-Viz : Semantic Relation Visualization in Chinese Wordnet

*Ming-Wei Xu, **Jia-Fei Hong,
***Shu-Kai Hsieh , *Chu-Ren Huang

* Institute of Linguistics Academia Sinica, Taiwan
** Graduate Institute of Linguistics, National Taiwan University
*** Department of English, National Taiwan Normal University

GWC-08, Szeged Hungary, Jan. 22-25, 2008

Outline

- Introduction
- Chinese Wordnet
- Applying visualization techniques to Lexical Semantic Relations
- Conclusion

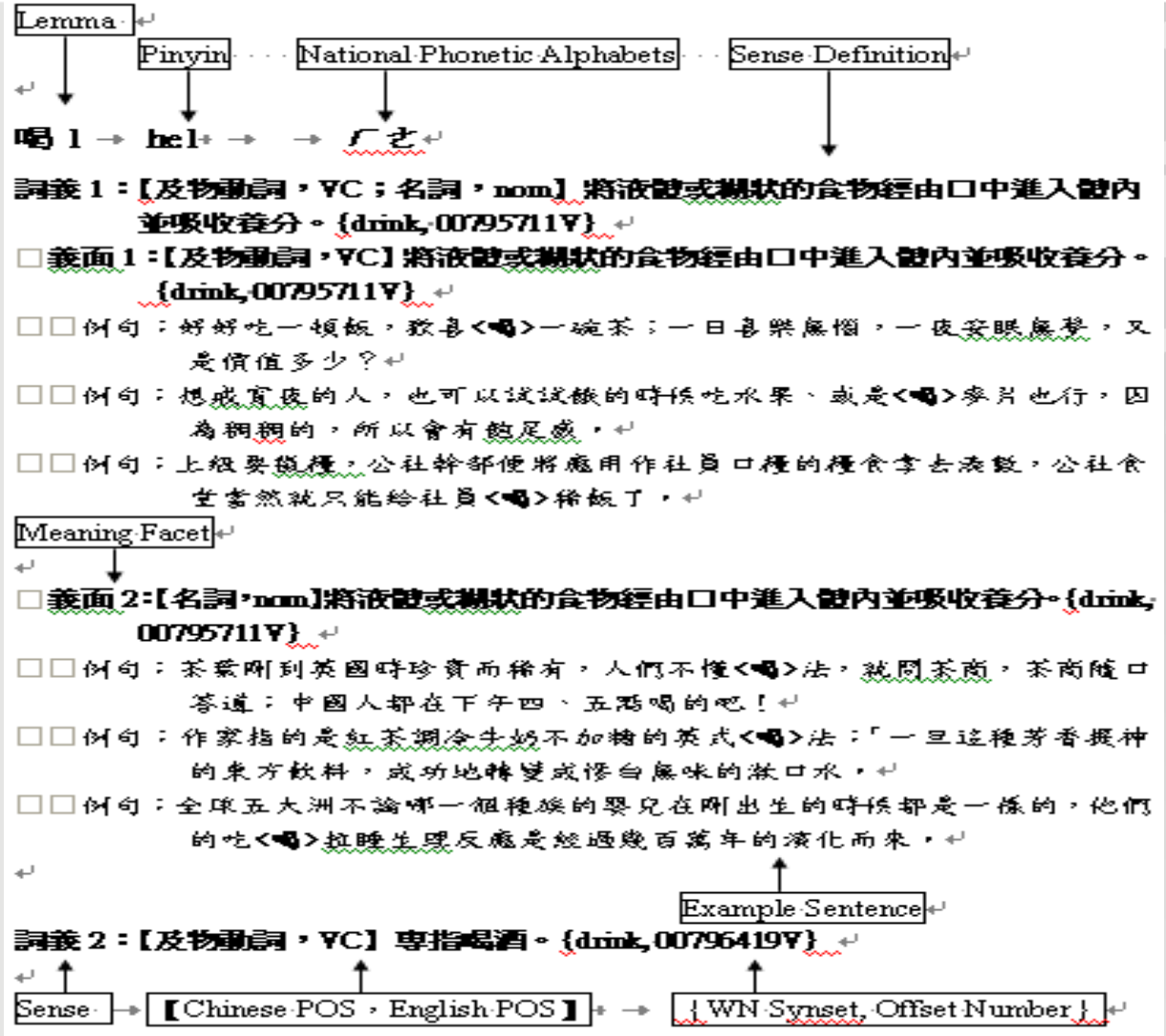
Introduction

- WordNet (Miller et al, 1990)
- Sinica BOW (Huang et al, 2004)
  - The Academia Sinica Bilingual Ontological Wordnet (Sinica BOW)
  - It integrates three resources: WordNet1.6/1.7.1, English-Chinese Translation Equivalents Database (ECTED), and SUMO.
- Chinese Wordnet at Academia Sinica

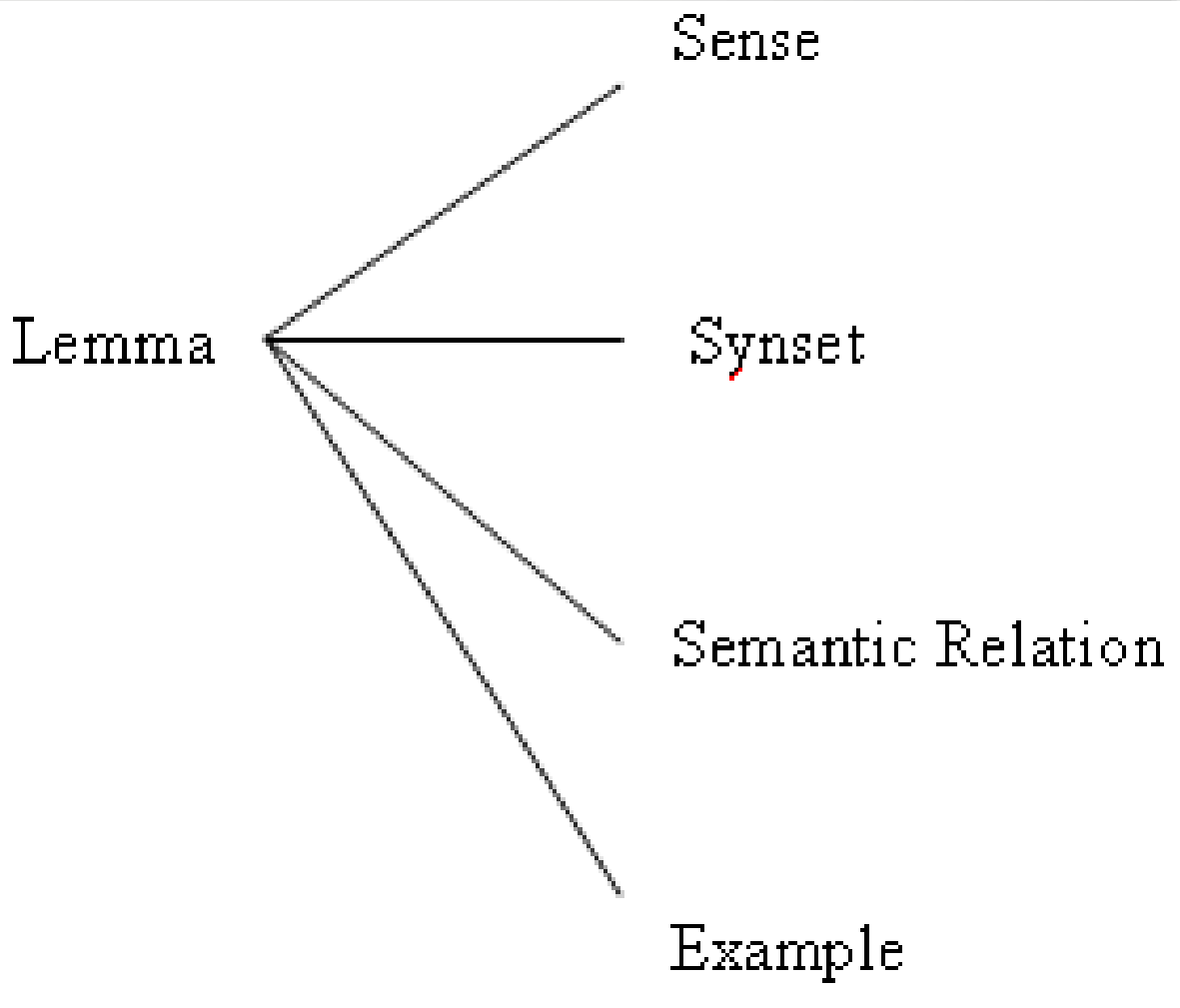
Chinese Wordnet

- CWN Group (Huang, 2003-)
- Highly linguistically motivated criteria for sense distinction
  - — Linguistic Felicity for sense/meaning facet
- Bilingual Ontological Lexical Resource
- LSRs bootstrapping from PWN/EWN

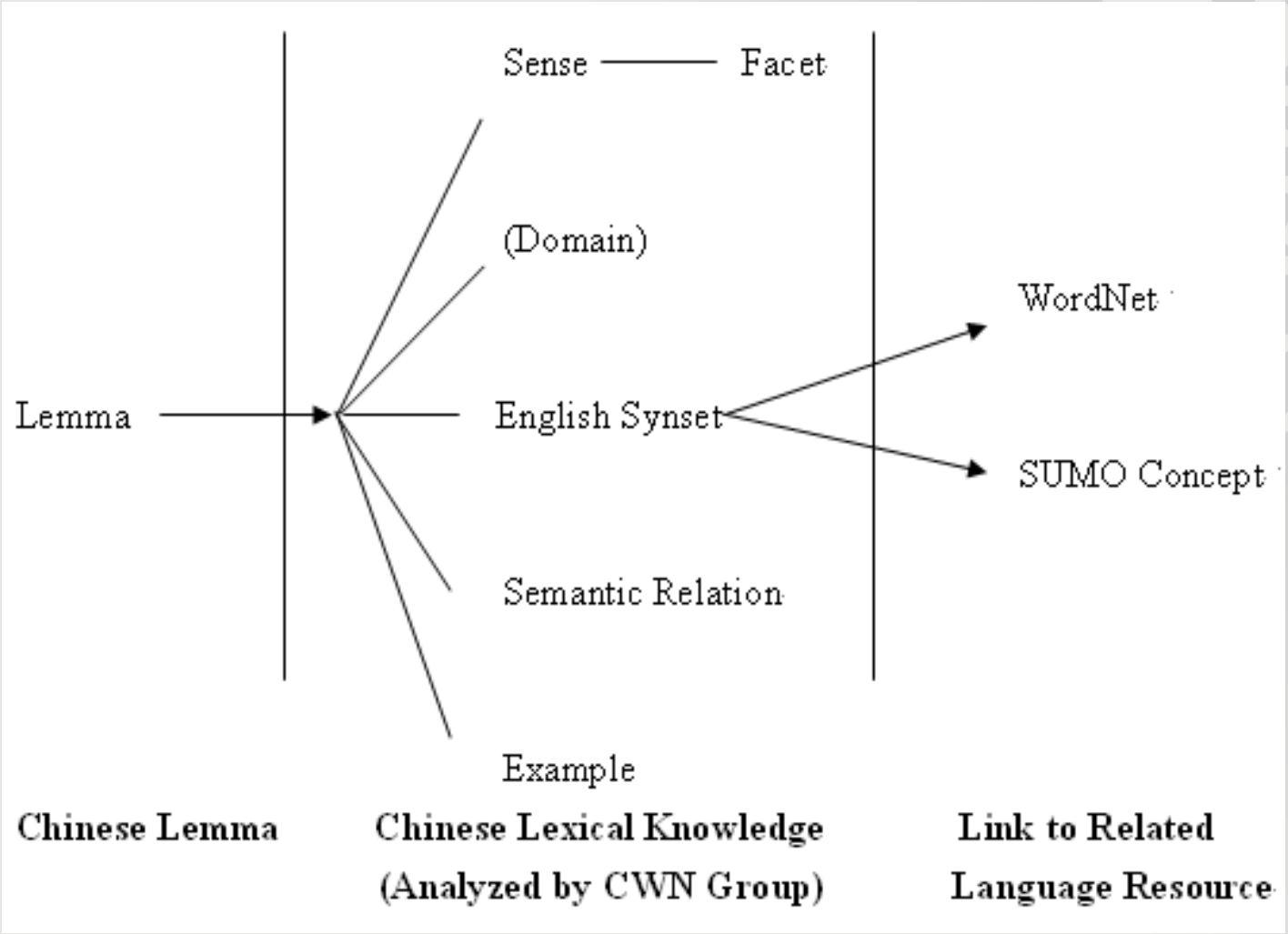
Example


Lemma
Pinyin
National Phonetic Alphabets
Sense Definition
喝 1 → he1 → → ㄏㄜ
詞義 1：[及物動詞，VC；名詞，nom] 將液體或糊狀的食物經由口中進入體內並吸收養分。{drink,00795711V}
義面 1：[及物動詞，VC] 將液體或糊狀的食物經由口中進入體內並吸收養分。{drink,00795711V}
例句：好好吃一頓飯，救急<喝>一碗茶；一日喜樂無憾，一夜安眠無夢，又是值值多少？
例句：想減肥夜的人，也可以試試餓的時候吃水果、或是<喝>麥片也行，因為稠稠的，所以會有飽足感。
例句：上級要獎懲，公社幹部便將應用作社員口糧的糧食拿去表數，公社食堂當然就只能給社員<喝>稀飯了。
Meaning Facet
義面 2：[名詞，nom] 將液體或糊狀的食物經由口中進入體內並吸收養分。{drink,00795711V}
例句：茶葉剛到英國時珍貴而稀有，人們不懂<喝>法，就問茶商，茶商隨口答道：中國人都在下午四、五點喝的吧！
例句：作家指的是紅茶調冷牛奶不加糖的英式<喝>法：「一旦這種芳香提神的東方飲料，成功地轉變成慘白無味的潑口水。
例句：全球五大洲不論哪一個種族的嬰兒在剛出生的時候都是一樣的，他們的吃<喝>拉睡生理反應是經過幾百萬年的演化而來。
Example Sentence
詞義 2：[及物動詞，VC] 專指喝酒。{drink,00796419V}
Sense
【Chinese POS，English POS】
{WN Synset, Offset Number}

Data Structure of WN



Data Structure of CWN

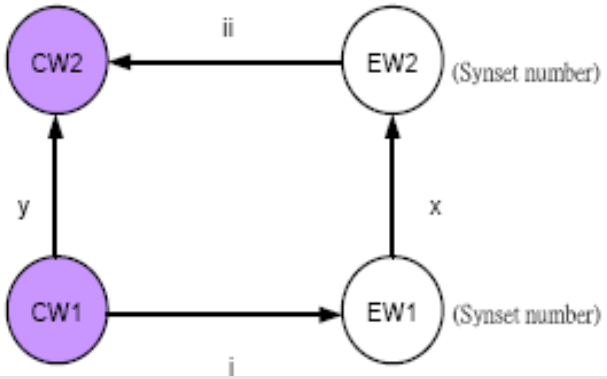


Bootstrapping Lexical Semantic Relations

- Cross-lingual conversion of Lexical Semantic Relations via inference rules (Huang et al. 2005; Hsieh et al. 2006)

Parallel Wordnets LSRs Conversion (complete model)

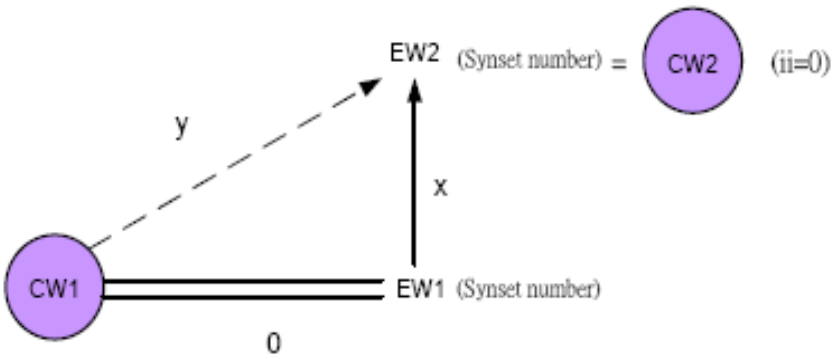
Translation-mediated LSR: The unknown LSR $y = i + x + ii$



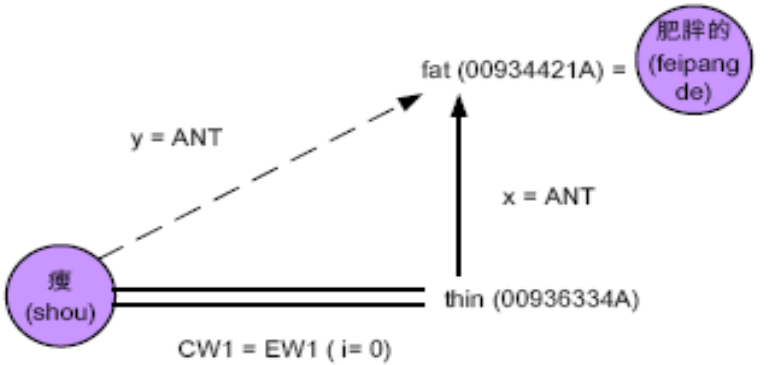
x = EW1 - EW2
y = CW1 - CW2
i = Translation LSR
ii = Translation LSR
The unknown LSR
y = i + x + ii

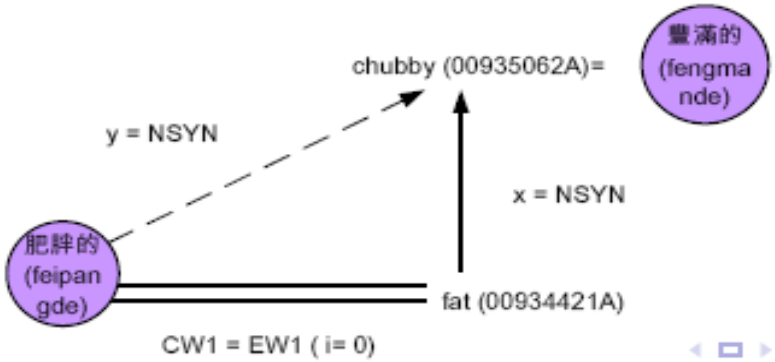
Parallel Wordnets LSRs Conversion (reduced model)

Translation-mediated LSR (When TEs are synonymous): The unknown LSR $y = 0 + x + 0 = x$ (for $i = 0$; $ii = 0$)

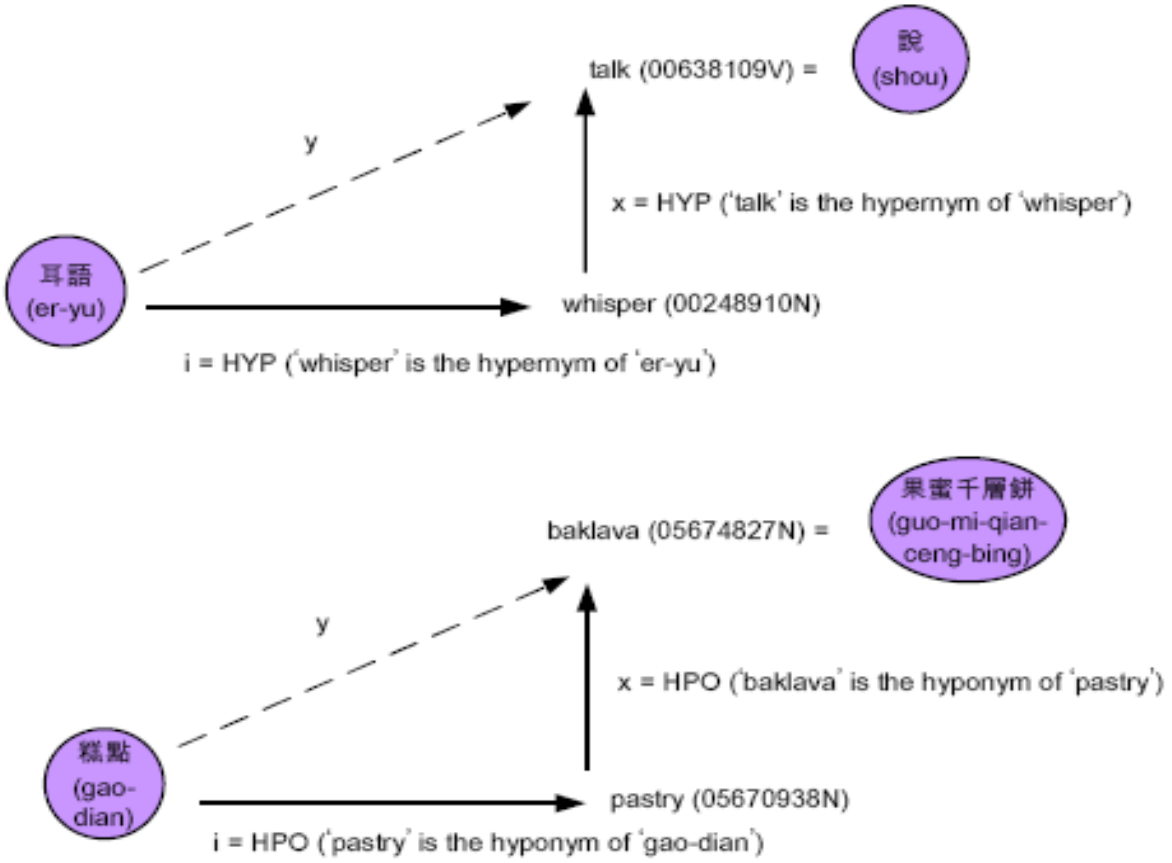


Example (When TEs are synonymous):





Example (When TEs are not synonymous):



	I	X	Y	Bootstrapped Results
1	HYP	ANT	ANT	{CW1, ANTONYM, CW2}
2	HYP	HYP	HYP	{CW1, HYPONOMY, CW2}
3	HYP	NSYN	HYP	{CW1, HYPONYM, CW2}
4	HYP	HOL	HOL	{CW1, HOLONYM, CW2}
5	HYP	all other LSRs	undecided	?
6	HPO	ANT	ANT	{CW1, ANTONYM, CW2}
7	HPO	HPO	HPO	{CW1, HYPONYM, CW2}
8	HPO	NSYN	HPO	{CW1, HYPONYM, CW2}
9	HPO	MER	MER	{CW1, MERONYM, CW2}
10	HPO	all other LSRs	undecided	?
11	NSYN	ANT	ANT	{CW1, ANTONYM, CW2}
12	NSYN	HYP	HYP	{CW1, HYPERNYM, CW2}
13	NSYN	HPO	HPO	{CW1, HYPONYM, CW2}
14	NSYN	NSYN	NSYN	{CW1, NEAR-SYNONYM, CW2}
15	NSYN	MER	MER	{CW1, MERONYM, CW2}
16	NSYN	HOL	HOL	{CW1, HOLONYM, CW2}
17	HOL	ANT	ANT	{CW1, ANTONYM, CW2}
18	HOL	HYP	HYP	{CW1, HYPONYM, CW2}
19	HOL	NSYN	HOL	{CW1, HOLONYM, CW2}
20	HOL	HOL	HOL	{CW1, HOLONYM, CW2}
21	HOL	all other LSRs	undecided	?
22	MER	ANT	ANT	{CW1, ANTOMY, CW2}
23	MER	HPO	HPO	{CW1, HYPONYM, CW2}
24	MER	NSYN	MER	{CW1, MERONYM, CW2}
25	MER	MER	MER	{CW1, MERONYM, CW2}
26	MER	all other LSRs	undecided	?

Current States

- From 2003 till September 2007,
  - 7198 lemma
  - 17932 senses
  - 4191 mapping to PWN synset (WN2.0:4134/WN2.1:30/WN3.0:27)
  - 13823 Synsets
  - 18006 Relations

Applying Visualization Technique to LSRs

- Previous works : WordNet TreeWalk (Bou, 2003); WordNet Connect (Fong, 2003); WordNet Relationship Browser (Alcock, 2004).

- More recently, Visual Thesaurus (ThinkMap, 2005); Visual WordNet Project (Kuo, 2005); WordNet Explorer (Collins, 2006)

Applying Visualization Technique to LSRs

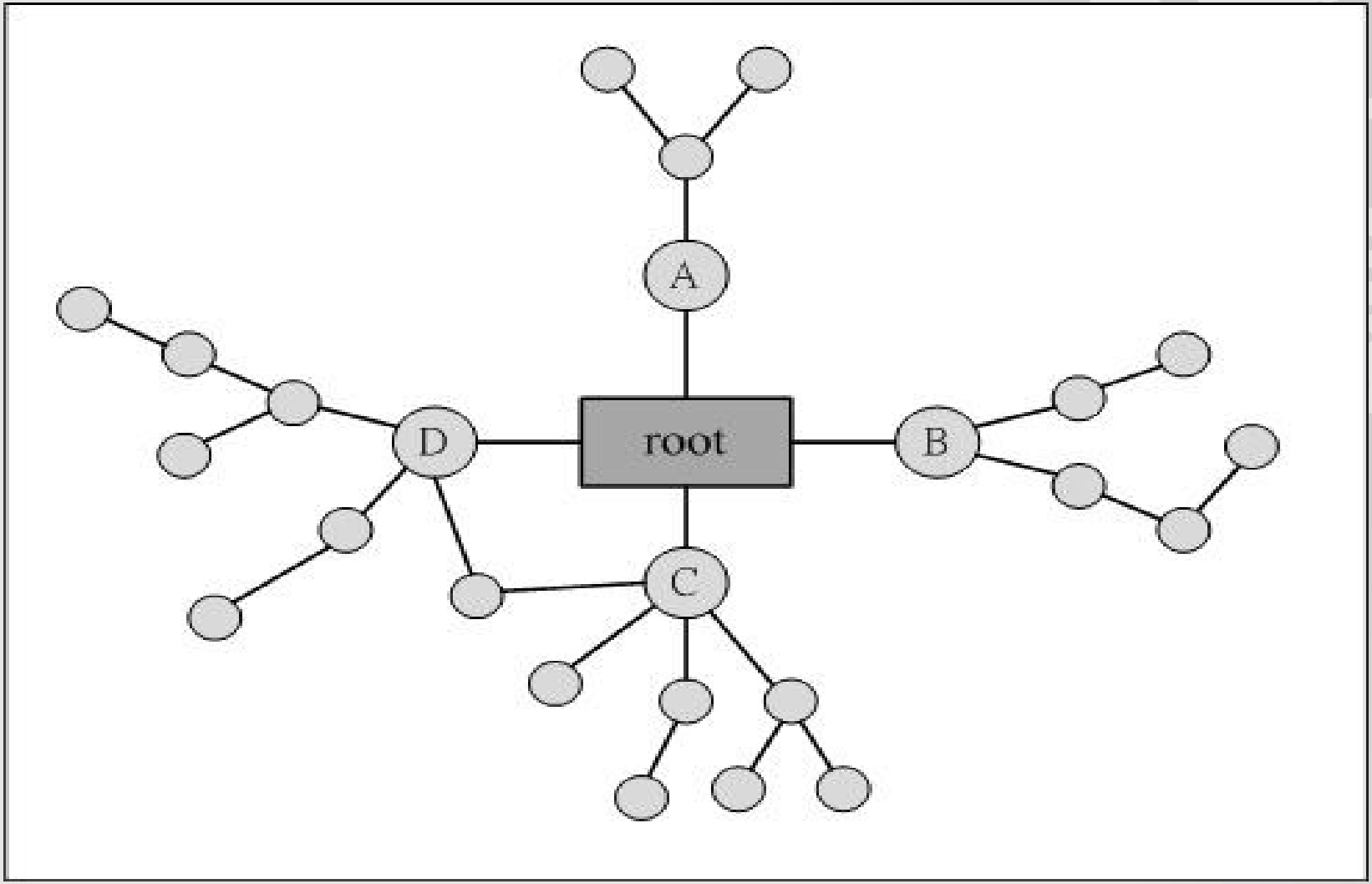
However, we need alternative tools to meet our needs:

- Not only show a full picture of WordNet relations, but also give context as well. (e.g., corpus, dictionary gloss, thesaurus, etc.)
- Distinctively show predicted LSRs with via bootstrapping rules (Huang et al, 2003) for the purpose of evaluation.
- Provide a window for showing concept clusters using morpho-semantic links

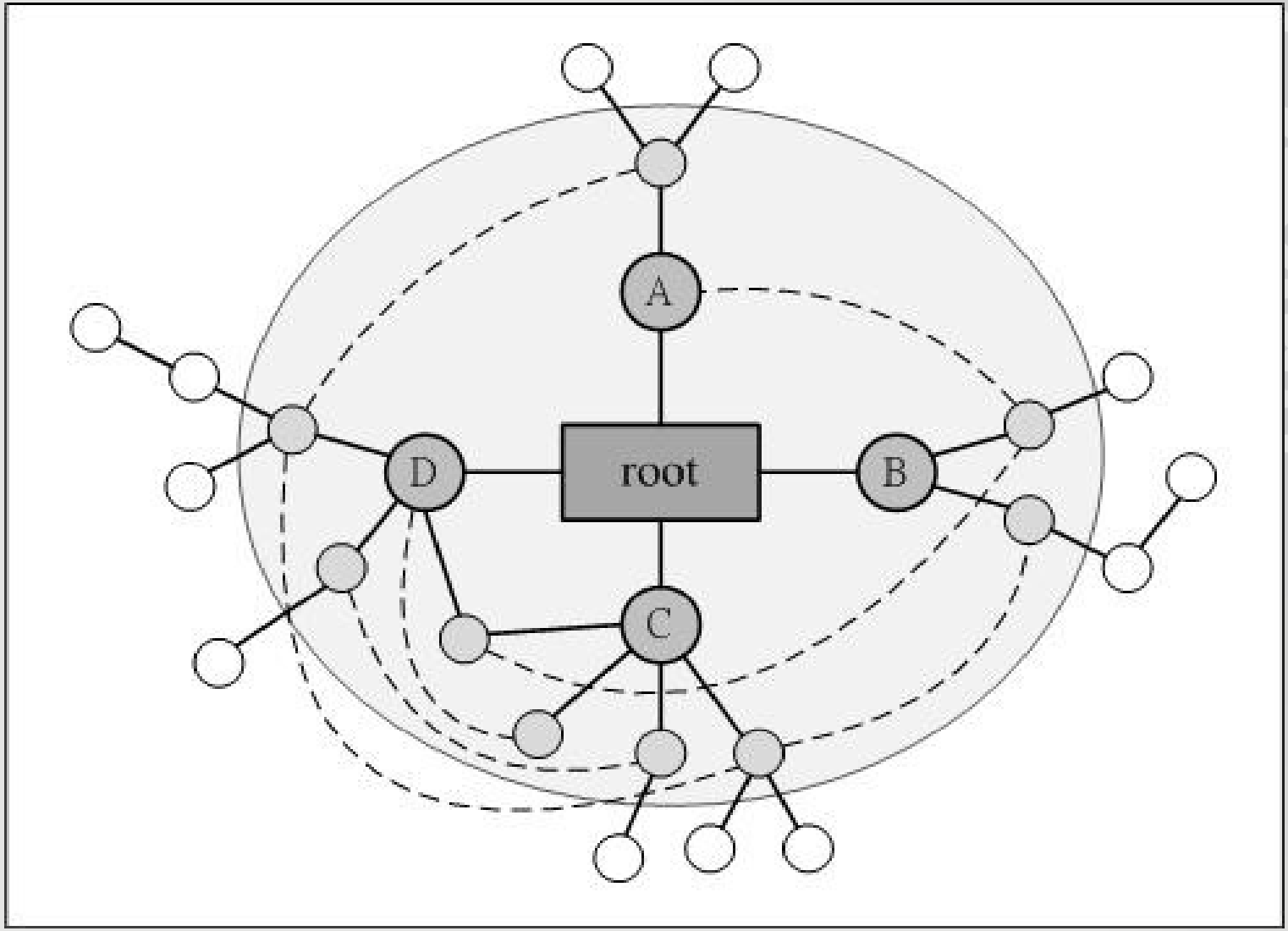
CWN-Viz: The First Try

- Interface for browsing integrated resources and evaluating bootstrapped LSRs.
- Technically, we follow the design paradigm based on "TouchGraph" (Google)- an open source graph layout system, to construct a working prototype of a visualization suite for Chinese Wordnet.
- For now, it can show all lemmas, senses, and semantic relations for a word form recorded in Chinese Wordnet, and basic measure of the the distance of each semantic relation.

The basic visualization construction



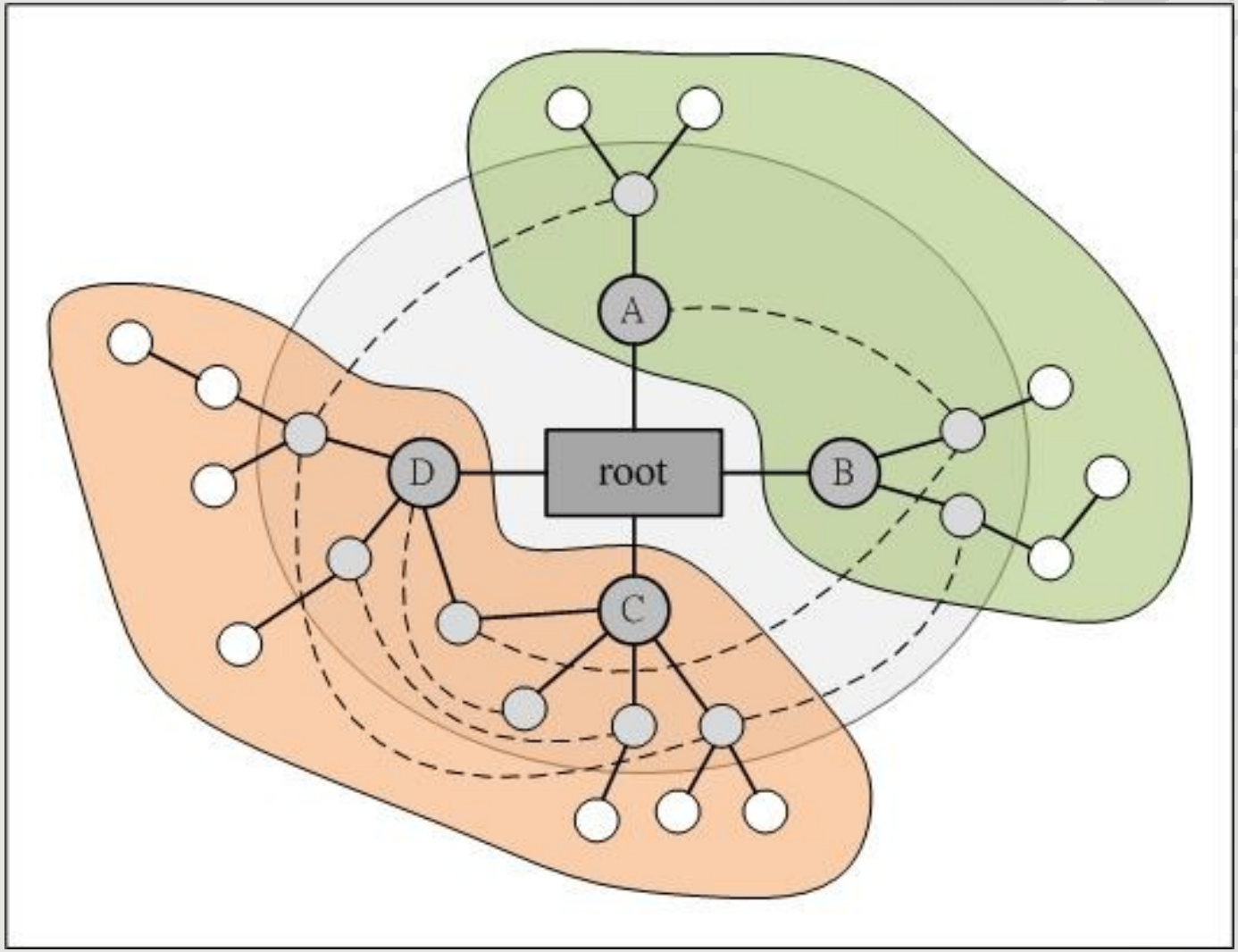
Visualization construction of Semantic relations



Calculating Principle

- A keyword root: a center node and extend two levels
- Based on the first level nodes to calculate sub-roots of the sub-trees
- These sub-trees
  - —Evaluate the relationship score for each sub-tree
- Calculate the relationship score for each sub-tree
  - —Present the calculating matrix for each cluster
- Select the most nodes of the numbers semantic relations until all nodes

The clusters --- Viz construction



正 2 → → → → The second lemma of 正

正 2(0620) → → The sixth senses of 正 2

美麗 mei3 li4 beautiful

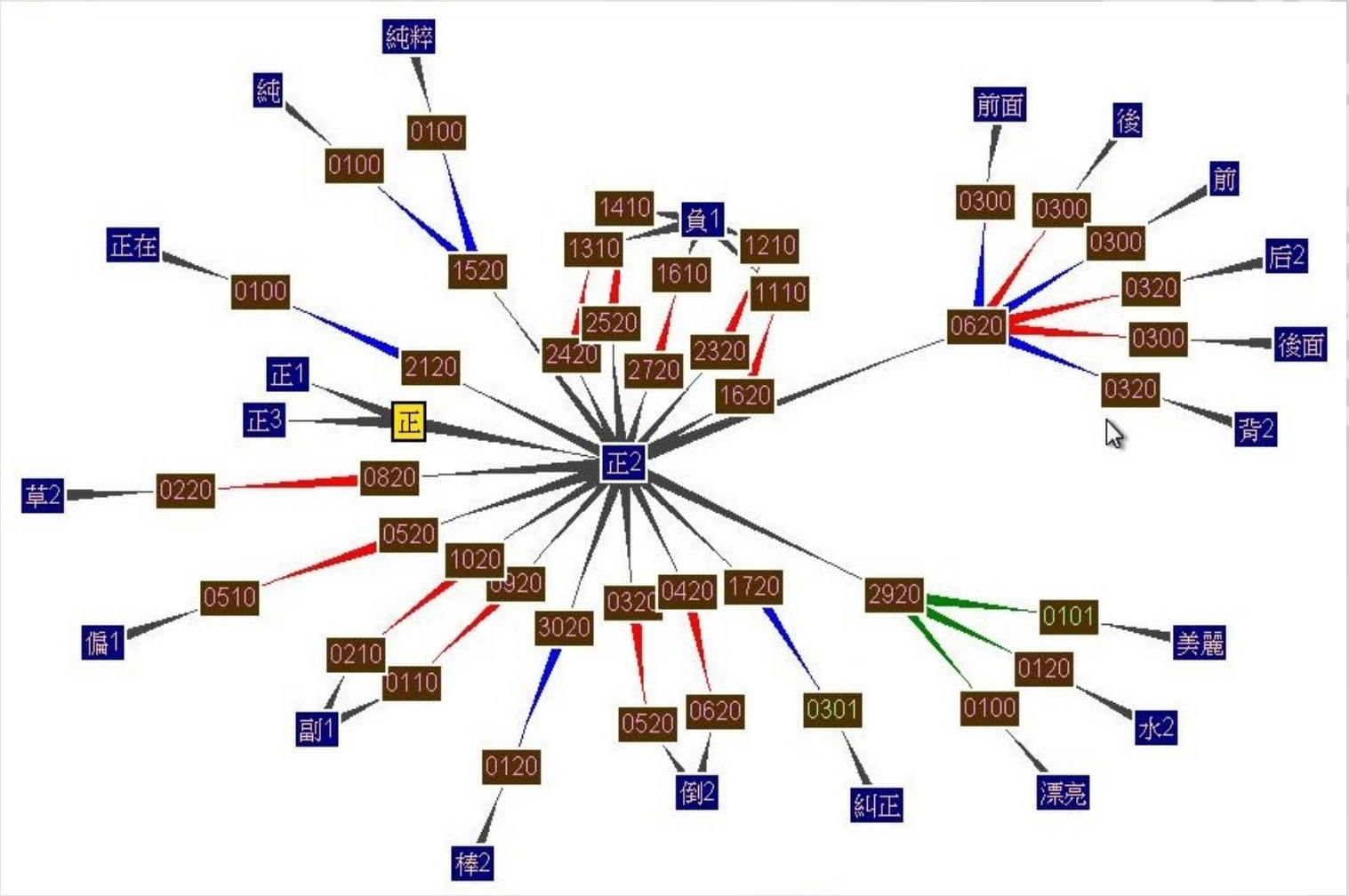
美麗(0101) → → The first meaning facet of the first sense of 美麗

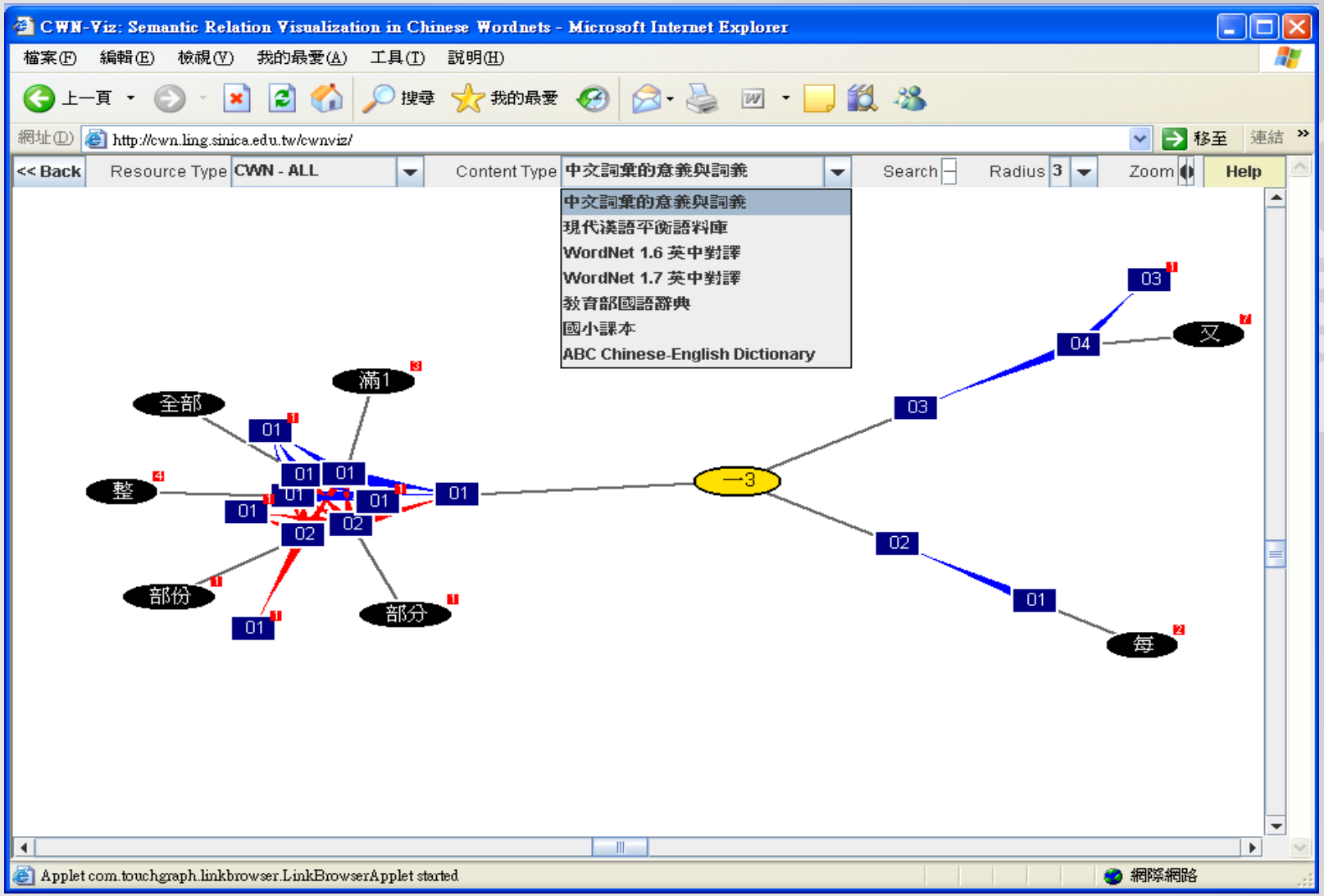
lemma : for lemma

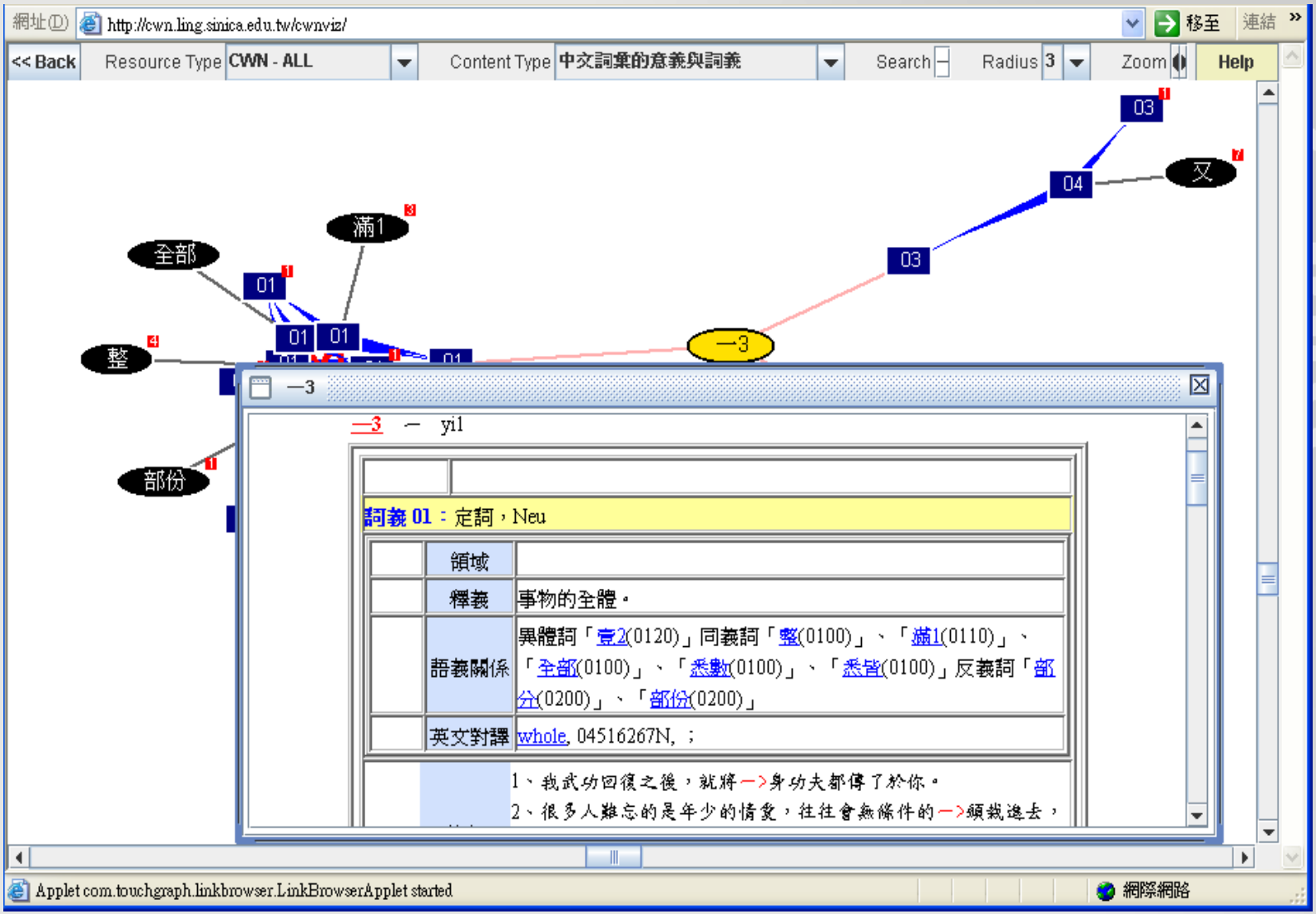
sense : for sense

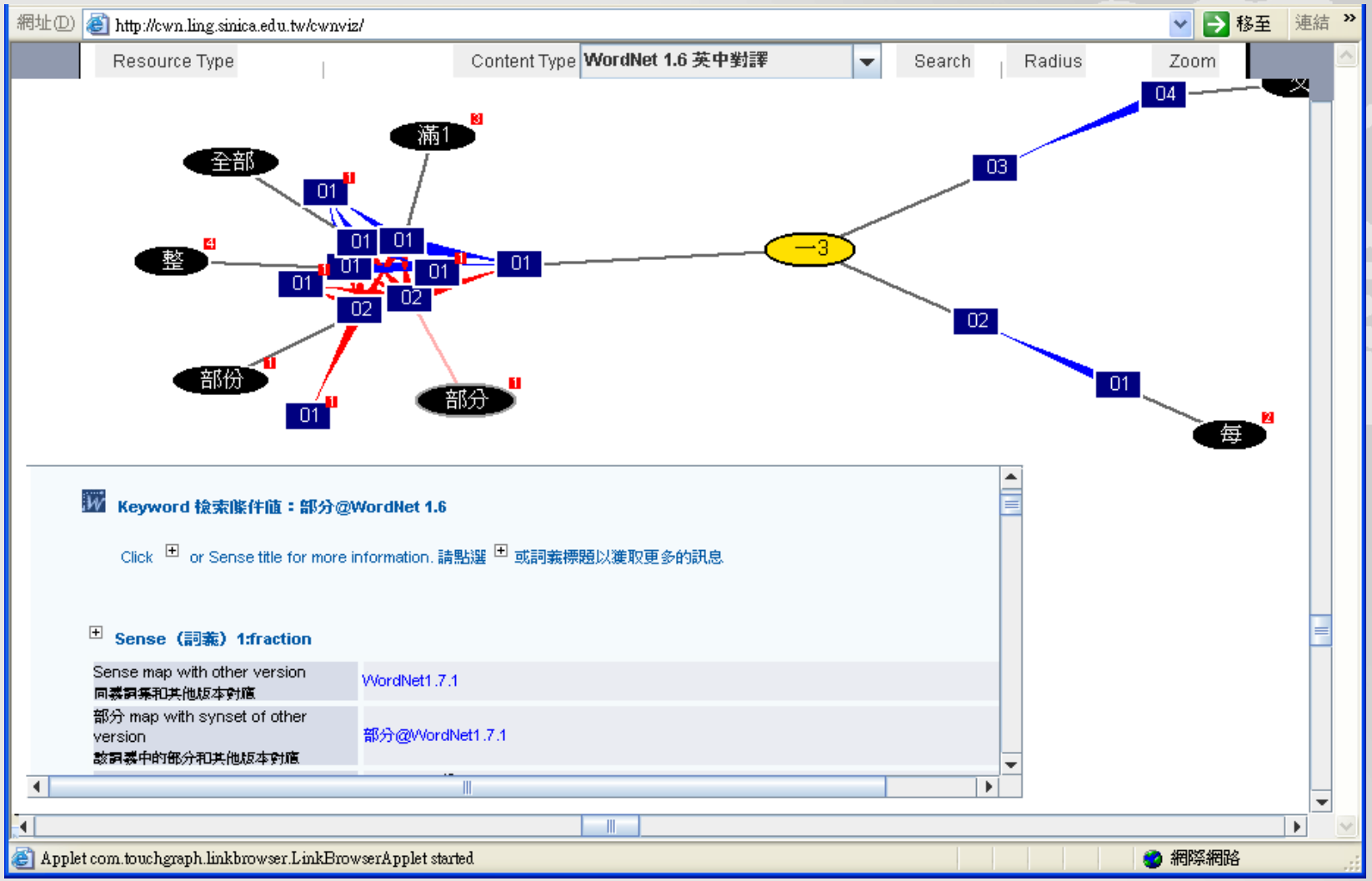
facet : for facet

Viz for 正 Zheng4 ('right')




CWN-Viz: Semantic Relation Visualization in Chinese Wordnets - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛
網址(D) http://cwn.ling.sinica.edu.tw/cwnviz/
移至 連結
<< Back
Resource Type CWN - ALL
Content Type 中文詞彙的意義與詞義
Search
Radius 3
Zoom
Help
中文詞彙的意義與詞義
現代漢語平衡語料庫
WordNet 1.6 英中對譯
WordNet 1.7 英中對譯
教育部國語辭典
國小課本
ABC Chinese-English Dictionary
一3
滿1
全部
整
部份
部分
每
又
Applet com.touchgraph.linkbrowser.LinkBrowserApplet started
網際網路


網址(D) http://cwn.ling.sinica.edu.tw/cwnviz/
移至
連結
<< Back
Resource Type
CWN - ALL
Content Type
中文詞彙的意義與詞義
Search
Radius
3
Zoom
Help
03
04
又
滿1
全部
01
01
01
03
整
一3
部份
一3
一 yi1
詞義 01：定詞，Neu
領域
釋義
事物的全體。
語義關係
異體詞「壹2(0120)」同義詞「整(0100)」、「滿1(0110)」、「全部(0100)」、「悉數(0100)」、「悉皆(0100)」反義詞「部分(0200)」、「部份(0200)」
英文對譯
whole, 04516267N, ;
1、我武功回復之後，就將一>身功夫都傳了於你。
2、很多人難忘的是年少的情愛，往往會無條件的一>頓栽進去，
Applet com.touchgraph.linkbrowser.LinkBrowserApplet started
網際網路


網址(D) http://cwn.ling.sinica.edu.tw/cwnviz/
移至
連結
Resource Type
Content Type
WordNet 1.6 英中對譯
Search
Radius
Zoom
全部
滿1
整
部份
部分
每
Keyword 檢索條件值：部分@WordNet 1.6
Click or Sense title for more information. 請點選 或詞義標題以獲取更多的訊息
Sense (詞義) 1:fraction
Sense map with other version
同義詞集和其他版本對應
WordNet1.7.1
部分 map with synset of other version
該詞義中的部分和其他版本對應
部分@WordNet1.7.1
Applet com.touchgraph.linkbrowser.LinkBrowserApplet started
網際網路

Conclusion

- Design the visualization tool such that various language resources and cross-lingual lexical semantic relations be processed.
- Facilitate linguists' use and understanding of Wordnets.

Thank You !

- Sinica BOW:

  http://bow.sinica.edu.tw
- Chinese Wordnet

  http://cwn.ling.sinica.edu.tw
- CWN-Viz Prototype Demo

  http://cwn.ling.sinica.edu.tw/cwnviz/