

Lessici Computazionali

Studio di un Caso: Wordnet

Roberto Basili

Department of Computer Science, System and Production
University of Roma, *Tor Vergata*
Via Della Ricerca Scientifica s.n.c., 00133, Roma, ITALY
e-mail: basili@info.uniroma2.it

Contenuti:

- Finalita' di un lessico computazionale
- WordNet: una rete semantica lessicale di ispirazione psicolinguistica
- I *Sensi* in Wordnet
- Relazioni lessicali e semantiche in Wordnet
- I verbi e i nomi in Wordnet
- EuroWordnet
- EuroWordnet in NAMIC

Finalita' di un lessico (computazionale) semantico :

- Fornire il "*referimento*" per la analisi semantica
- Associare alle entita' *proprieta' linguistiche* ed *extra-linguistiche*
- Organizzare le entita' in *strutture relazionali*
- Rappresentare le *relazioni semantiche*
- Fondare una *organizzazione concettuale* del mondo e ...
- determinare la *ontologia* di riferimento
- Supportare inferenze "*tipiche*" (e.g. IS_A")

Wordnet

- Nasce da una iniziativa del Cognitive Science Laboratory della Università' di Princeton, sotto la guida del prof. George Miller (1987)
- Scopo:
 - Fornire una descrizione a partire da principi *lessicografici*
 - Rappresentare forme di associazione psicologicamente fondate
 - Raggiungere una dimensione realistica (*sensi e parole*)

Wordnet

- Il sistema e' organizzato in strutture di grafi indipendenti per ogni categoria grammaticale (POS)
- Scala
 - Nomi: 109,195
 - Verbi: 11,088
 - Aggettivi: 21,460
 - Avverbi: 4,607

I Sensi in Wordnet :

- L'unità fondamentale della organizzazione lessicale di Wordnet e' il *synonymy set*, o *synset*
- Un *synset* e' formato da un insieme di parole che (secondo un certo aspetto del loro significato) sono *sinonimi*

1. set, circle, band, lot -- (an unofficial association of people or groups;
"the smart set goes there"; ...)
2. band -- (instrumentalists not including string players)
3. band, stria, striation -- (a stripe of contrasting color;
"chromosomes exhibit characteristic bands")
4. band, banding, stripe -- (a strip or stripe of a contrasting color or material)
5. dance band, band, dance orchestra -- (a group of musicians playing popular
music for dancing)
6. band -- (a range of frequencies between two limits)
7. band -- (something elongated that is worn around the body
or one of the limbs)
8. ring, band -- (jewelry consisting of a circular band of a precious metal
worn on the finger; "she had rings on every finger")
9. band -- (put around something to hold it together)

L'organizzazione dei Sensi in Wordnet :

- I *sysnet* sono organizzati in Wordnet secondo relazioni
 - lessicali: e.g. *sinonimia*, *antinomia*
 - semantico/concettuali: *hyponymy*, *hyponimy*

*** HYPONIMY: is_a_kind_of "band"

2. band -- (instrumentalists not including string players)

=> brass band -- (a group of musicians playing only brass
and percussion instruments)

=> concert band, military band -- (a group of musicians playing
brass and woodwind and percussion instruments)

*** HYPERONIMY: "band" is_a_Kind_of

2. band -- (instrumentalists not including string players)

=> musical organization, musical organisation, musical group
-- (an organization of musicians who perform together)

=> organization, organisation -- (a group of people who work together)

=> social group -- (people sharing some social relation)

=> group, grouping -- (any number of entities
(members) considered as a unit)

L'organizzazione dei Sensi in Wordnet (2):

- I *sysnet* sono descritti da esempi d'uso o "glosse"

*** GLOSS

2. bad, big -- (very intense; "a bad headache"; "in a big rage";
"had a big (or bad) shock"; "a bad earthquake"; "a bad storm")

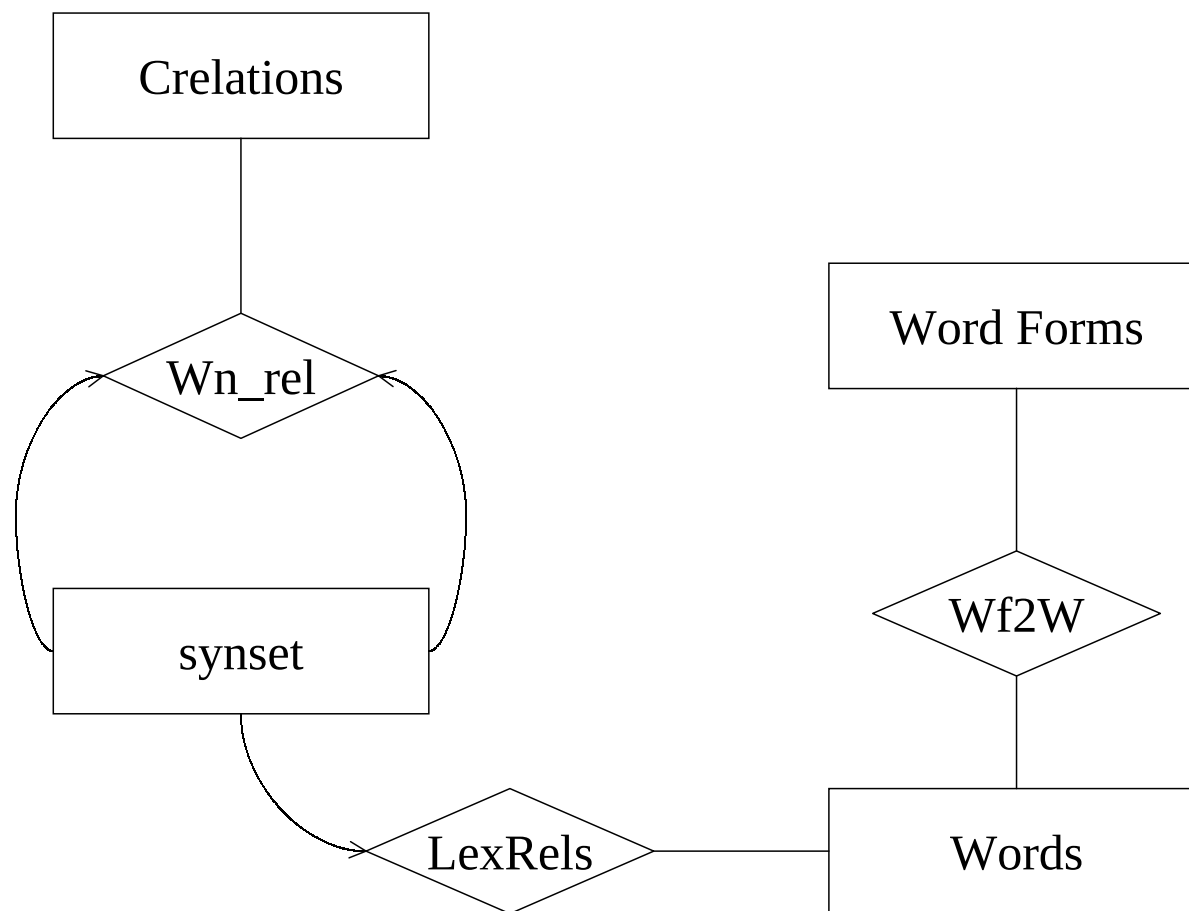
*** ANTONIMY: is_a_contrary_of "bad"

bad, big -- (very intense; "a bad headache"; ...)

INDIRECT (VIA intense) -> mild --

(moderate in type or degree or effect or force;
far from extreme; "a mild winter storm"; "a mild fever";
"fortunately the pain was mild"; "a mild rebuke";
"mild criticism")

Organizzazione dei sensi in WN:



Wordforms: scanning:

The noun "scanning" has 1 sense in WordNet.

1. scanning -- (the process of translating photographs into a digital form that can be recognized by a computer)

The verb "scan" has 8 senses in WordNet.

1. scan -- (examine minutely or intensely)
2. scan, skim, rake, glance over, run down -- (examine hastily)
3. scan -- (glance over or read superficially)
4. scan -- (make a wide, sweeping search of; "The beams scanned the night sky")
5. scan -- (conform to a metrical pattern; of poetic verse)
6. scan -- (move a light beam over; in electronics, to reproduce an image)
7. scan -- (read metrically; "scan verses")
8. read, scan -- (obtain data from magnetic tapes; "This dictionary can be read by the computer")

Wordnet (1.7): Scala:

POS	Unique Strings	Synsets	(W,Sense) Pairs
Noun	109,195	75,804	134,716
Verb	11,088	13,214	24,169
Adjective	21,460	18,576	31,184
Adverb	4,607	3,629	5,748
Totals	146,350	111,223	195,817

Wordnet (1.7): Polysemy information:

POS	Monosemous Words and Senses	Polysemous Words	Polysemous Senses
Noun	94,685	14,510	40,002
Verb	5,920	5,168	18,221
Adjective	15,981	5,479	15,175
Adverb	3,820	787	1,900
Totals	120,406	25,944	75,298

Wordnet: Average polisemy (AvPol) :

POS	Including Monosemous Words	Excluding Monosemous Words
Noun	1.23	2.75
Verb	2.17	3.52
Adjective	1.45	2.76
Adverb	1.24	2

Topmost synsets:

The noun "bank" has 10 senses in WordNet.

1. depository financial institution, bank, banking concern, banking company
-- (a financial institution that accepts deposits and
channels the money into lending activities; ...
2. bank -- (sloping land (especially the slope beside a body of water); ...

Sense 1

depository financial institution, ...

=> financial institution, financial organization, financial organisation ...

=> institution, establishment ..

=> organization, organisation ...

=> social group -- ...

=> group, grouping ...

Sense 2

bank -- (sloping land (especially the slope beside ...

=> slope, incline, side ...

=> geological formation, geology, formation -- ...

=> natural object -- ...

=> object, physical object ...

=> entity, physical thing ...

Topmost synsets (2):

- **entity, physical thing** (that which is perceived or known or inferred to have its own physical existence (living or nonliving))
- **psychological_feature**, (a feature of the mental life of a living organism)
- **abstraction**, (a general concept formed by extracting common features from specific examples)
- **state**, (the way something is with respect to its main attributes; ..
- **event**, (something that happens at a given place and time)
- **act, human_action, human_activity**, (something that people do or cause to happen)
- **group, grouping**, (any number of entities (members) considered as a unit)
- **possession**, (anything owned or possessed)
- **phenomenon**, (any state or process known through the senses rather than by intuition or reasoning)

Informazione grammaticale in Wordnet:

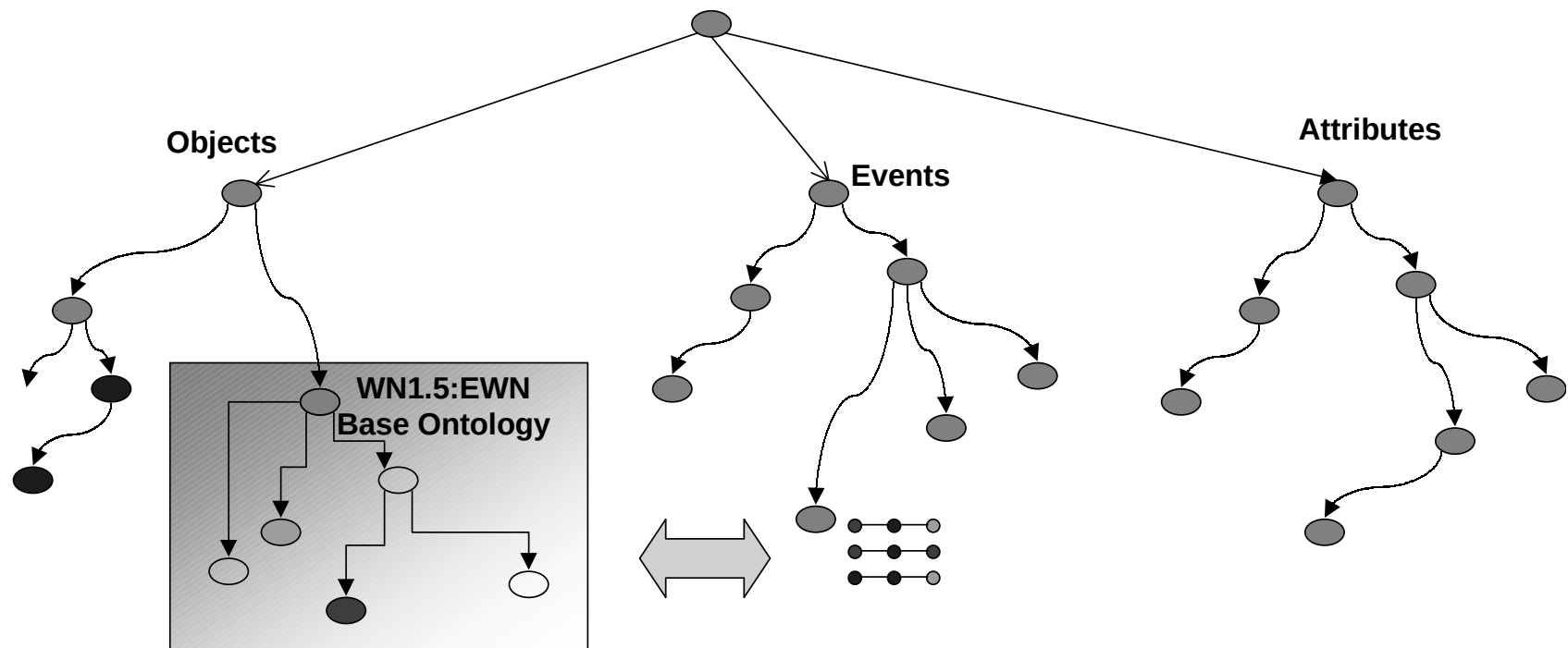
Sense 2

yield, give, afford -- (be the cause or source of;
"He gave me a lot of trouble";
"Our meeting afforded much interesting information")
*> Something ----s something

Sense 3

give -- (transfer possession of something concrete or abstract to somebody;
"I gave her my money";
"can you give me lessons?";
"She gave the children lots of love and tender loving care")
*> Somebody ----s somebody something
*> Somebody ----s something to somebody

Wordnet in NAMIC:



Ontologia in NAMIC:

```
/* NAMIC TOP ONTOLOGY USING EWN WITH ILI WN1.6 */
```

```
....
```

```
object(X) ==>
```

```
    text(X) v
```

```
    animate(X) v
```

```
    inanimate(X) v
```

```
    date(X) v /* Named Entity date recognized by the parser */
```

```
    action(X) v
```

```
    abstraction(X) v
```

```
    attribute(X) v
```

```
    wn_person(X) v
```

```
    wn_quantity(X) v
```

```
    time(X).
```

```
...
```

Ontologia in NAMIC:

Integrazione di (E)WN

```
inanimate(X) ==>
    location(X) v /* NE location recognized by the parser */
    wn_inanimate(X). /* Common nouns objects (from Eurowordnet) */

wn_inanimate(X) ==>
    ewn9457(X). /* object_1 */

action(X) ==>
    ewn17487(X) v /* act_1 */
    ewn19671(X). /* communication_1 */

abstraction(X) ==>
    ewn13243(X) v /* cognition_1 */
    ewn18916(X). /* relation_1 */
```

Ontologia in NAMIC: Eventi:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%  finance first level
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
finance_event(X) ==>
```

```
    company_relationships(X) v
    industrial_activities(X) v
    governmental_activities(X) v
    job_market(X) v
    market_industrial_sector(X) v
    stock_market(X).
```

```
...
```

```
market_industrial_sector(X) ==>
```

```
    '5-1'(X) v '%Position vs. Competitors Events'
    '5-3'(X).  '%Market Sector Events'
```

Ontologia in NAMIC - IE patterns:

```
'5-3'(X) ==> '5-3_0_23'(X).
'5-3_0_23'(X) ==> deliver(X) v draw(X) v embrace(X) v hit(X) v
                  improve(X) v offer(X) v provide(X) v raise(X) v reach(X) v
                  receive(X) v remove(X) v supply(X) v welcome(X).

/* -----
ontology('5-3','Market Sector Events','5-3_0_23',
[deliver,draw,embrace,hit,improve,offer,
 provide,raise,reach,receive,remove,supply,welcome]).
----- */
props( '5-3_0_23'(E), [
    ( consequence(E,[namic_relation(E,'5-3'),
                      namic_object(E,Obj),
                      namic_agent(E,Subj)]) :-

        nodeprop(E,lobj(E,Obj)),
        Obj <- ewn13243(_)          /* cognition knowledge */ ,
        nodeprop(E,lsubj(E,Subj)),
        Subj <- ewn261466(_)        /* activity */
    )
]).
```