

WORDNET

An on-line lexical DataBase

LOMBARDI NICOLA

Corso di Elaborazione Del Linguaggio Naturale

Laurea Specialistica in Informatica
Università di Pisa – Anno Accademico 2005-2006

1

INTRODUZIONE

Dizionari Tradizionali

- Realizzati su carta stampata
- Ordinati alfabeticamente
- Ricerca spesso noiosa e lunga (rispetto alla possibilità di un dizionario elettronico)
- Alcuni tipi di ricerca risultano praticamente impossibili (da un punto di vista temporale)
 - **Esempio:**
Pino: Conifera sempreverde...
 - Se volessimo trovare altre *Conifere* saremmo costretti a sfogliare tutto il dizionario dalla A alla Z

2

INTRODUZIONE

Obiettivi

- Strumento che sfrutti nuovi sistemi tecnologici
 - Computer
 - Internet
- Miglioramenti
 - Velocità
 - Praticità
 - Utilizzo
 - Efficienza
 - Etc...
- Realizzazione di un dizionario ordinato concettualmente e non alfabeticamente
- Realizzazione WordNet
 - Princeton University (New Jersey) – 1985
 - Prof. George A. Miller

3

INTRODUZIONE

Idea Base

- Divisione delle parole in categorie
 - Nomi
 - Aggettivi
 - Verbi
 - Avverbi
- Organizzazione basata sulla semantica delle parole

4

INTRODUZIONE

Matrice Lessicale

- Una parola è un'associazione tra una stringa sintattica ed un concetto
- Matrice lessicale
 - Stringhe (colonne)
 - Concetti (riga)

5

INTRODUZIONE

Matrice Lessicale (2)

- Difficoltà nel rappresentare concetti semantici
- Insieme di stringhe con cui tali concetti possono essere rappresentati
- Organizzazione di gruppi di sinonimi: SYNSET
- Una parola può comparire in due gruppi diversi
 - { board 1, plank } ; { board 2, committee }
- Talvolta è comunque presente una piccola descrizione

6

INTRODUZIONE

Relazioni tra Synsets

- Si possono creare relazioni semantiche tra Synset
- Se c'è una relazione semantica R tra da un Sysnet $X = \{x, x', x'', \dots\}$ a $Y = \{y, y', y'', \dots\}$ allora esiste anche una relazione R' da Y a X
- Possibilità di passare per mezzo delle relazioni da un Synset all'altro
- Relazioni più importanti:
 - Sinonimia
 - Meronimia
 - Iperonimia/Iponimia
 - Contrarietà

7

NOMI

Dizionari tradizionali

- Nella definizione di una parola compare:
 - Super-ordinato
 - Caratteristiche salienti
 - Esempio:
 - Albero:
 - Pianta
 - fusto ritto, legnoso che nella parte superiore si spande in rami ...
- Non vengono citate alcune caratteristiche più intrinseche
 - Radici, organismo vivente ecc...
 - Definizione di pianta
- Presuppongono alcune conoscenze di base
- Se volessimo trovare altri tipi di pianta o i diversi tipi di albero dovremmo scorrere tutto il dizionario

8

NOMI

Formazione Synset di nomi

- Formazione di insiemi (SYNSET) di sinonimi
 - {bag, case, suitcase....}
- In caso di polisemia una parola può apparire in due Synset differenti: uno per ogni diverso significato che la parola può assumere
 - {case1, container,...}
- Talvolta, causa mancanza di sinonimi, vengono inseriti, racchiusi tra parentesi tonde all'interno del Synset, brevi commenti esemplificativi sul significato di una parola

9

NOMI

Scala gerarchica (1)

- Struttura ad albero per mezzo dei rapporti di iponimia/iperonimia
- Concetti molto generali alla radice
- Elementi specifici scendendo verso i rami
- L'operatore "@->" fa salire di livello nella scala gerarchica: fornisce quindi l'immediato iperonimo.
 - quercia @-> albero @-> pianta @-> organismo
- Al contrario l'operatore "->" fornisce l'immediato iponimo

10

NOMI

Scala gerarchica (2)

- I Synset vengono disposti ad albero e collegati per mezzo di puntatori
 - Ogni Synset avrà al proprio interno un puntatore verso il diretto iperonimo e un puntatore per ogni suo diretto iponimo
 - Esempio :
{ mammifero, animale@ , tigre -, gatto -, ... }
- Risposta al problema di trovare altri elementi appartenenti ad una stessa categoria
 - Da "Pino" si risale a "Conifera" e si riscalda a tutti i tipi di Conifera (lucido 2)
- Rispecchia organizzazione mentale

11

NOMI

Scala gerarchica (3)

- Si pone il problema di stabilire i Synset più generali che stanno alla radice delle scale gerarchiche.
- Inizialmente venne assunta una unica categoria (Synset) con due diretti iponimi
 - { entità }
 - { oggetti, cose }
 - { idee }
- Tale scelta è stata rivista e attualmente i Synset iniziali sono 25 suddivisi per differenti campi tematici

12

NOMI

Scala gerarchica (4)

- Lista Synset iniziali

{ Act, action, activity}	(natural object)
{ animal, fauna}	(natural phenomenon)
{ artifact}	(person, human being)
{ attribute, property}	(plant, flora)
{ body, corpus}	(possession)
{ cognition, knowledge}	(process)
{ communication}	{ quantity, amount}
{ event, happening}	(relation)
{ feeling, emotion}	(shape)
{ food}	(state, condition)
{ group, collection}	(substance)
{ location, place}	(time)
{ motive}	

13

NOMI

Scala gerarchica (5)

- In realtà sono state poi inseriti due Synset iperonimi dei 25 precedenti
 - {living thing, organism}
 - {non-living thing, object}
- A sua volta questi Synset sono stati raggruppati in
 - {thing, entity}

14

NOMI

Meronymia

- Un oggetto A è meronimo di B se A è parte fisica di B
 - *Ruota* è meronimo di *Automobile*
- Asimmetrica, transitiva
- La relazione inversa si chiama Olonimia
- La Meronymia si mantiene nei rapporti di iponimia:
 - Se A è meronimo di B e C è iponimo di B allora A è anche meronimo di C
 - *Ruota* è meronimo sia di *automobile* che di *fuoristrada*

15

NOMI

Meronymia (2)

- I rapporti di meronymia sono implementati in WordNet
- All'interno di ogni Synset viene indicato se il gruppo di parole che lo costituiscono sono meronime di un altro concetto espresso da un diverso Synset
- Vengono distinti tre tipi di meronymia
 - Component-object (ramo, albero) indicata con #p→
 - Member – collection (albero, foresta) indicata con #m→
 - Stuff – object (alluminio, aeroplano) indicata con #s →

16

NOMI

Meronomia (3)

- Sfruttando la transitività della meronomia si corre il rischio di arrivare a conclusioni che non rispecchiano il senso comune.
 - *Maniglia* è meronimo di *porta*, *porta* è meronimo di *casa* → *maniglia* è meronimo di *casa*
- Estremizzando tale concetto
 - *Atomo* è meronimo di ogni cosa
- E' stato deciso in maniera arbitraria di togliere i collegamenti di meronomia nel momento in cui A meronimo di B non rappresenta più una caratteristica saliente e distintiva di B.

17

NOMI

Contrarietà

- Per ogni insieme di termini viene infine inserito un collegamento al Synset che corrisponde al suo esatto contrario per mezzo dell'operatore "!→"
 - {man, woman !, person @,(a male person)}
- La contrarietà è ovviamente simmetrica per definizione
 - {woman, man !,}

18

NOMI

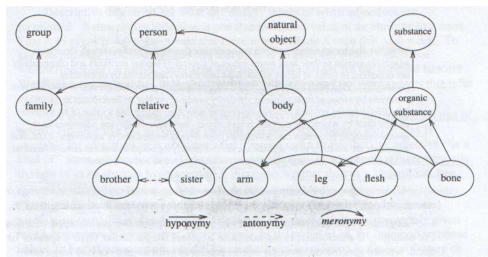
Riassumendo

- I nomi sono raggruppati in insiemi (Synset) di parole sinonime
- In ogni Synset { } possono inoltre inseriti brevi commenti racchiusi tra parentesi tonde e collegamenti a :
 - @ iperonimo
 - ~ iponimi
 - ! contrari
 - # meronimo

19

NOMI

Schema esemplificativo



20

AGGETTIVI

Aspetti generali

- Sono divisi in due classi:
 - Descrittivi
 - Relazionali
- Descrittivi
 - sono utilizzati per attribuire un valore ad un sostantivo
 - Caratterizzati da frequente bipolarismo (nella maggior parte dei casi ad ogni nome è associato un suo contrario)
- Relazionali
 - Accostati ad un sostantivo ne modificano il significato

21

AGGETTIVI DESCRITTIVI

- Un 'aggettivo e' descrittivo se serve ad aggiungere una qualità o un valore ad un sostantivo [Zingarelli]
- Visto in termini matematici può essere paragonato ad una funzione $A(x) = \text{Agg.}$
 - A è il sostantivo riferito all'aggettivo da attribuire all'oggetto x
 - Esempio
A = pesantezza; x = valigia; $A(x) = \text{pesante}$
- La relazione di base è la contrarietà
 - Simmetrica
 - Coinvolge un gran numero di aggettivi
 - Si rappresenta "!" →"

22

AGGETTIVI DESCRITTIVI

Organizzazione

- L'organizzazione tra i Synset si basa proprio sul rapporto di contrarietà tra aggettivi
- Differente da quella dei nomi che si basava su iperonimia/iponima
- Non hanno un struttura gerarchica
- Organizzazione in cluster

23

AGGETTIVI DESCRITTIVI

Problemi sulla contrarietà

- Se due aggettivi hanno un significato simile tra di loro perché non hanno lo stesso contrario?
 - Esempio
Heavy e *Weighty* hanno come contrari *light* e *weightless*
- Perché alcuni aggettivi non hanno un contrario?
 - Esempio
Ponderous non ha contrario.
Potremmo pensare a *light* ma il contrario di *light* è *heavy* (si perde la simmetria)

24

AGGETTIVI DESCRITTIVI

Struttura a cluster

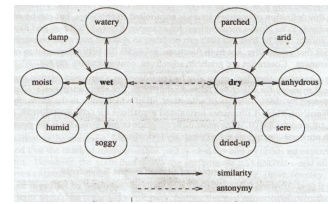
- Una struttura a gruppi di Sinonimi come i nomi non è adatta in quanto perderemmo i legami tra aggettivi contrari
- Vogliamo fare in maniera di non perdere anche legami tra aggettivi simili
- Il risultato è una struttura a cluster
- Per ogni coppia (a_1 , a_2) di aggettivi tra loro contrari vengono realizzati due cluster in ognuno dei quali:
 - la coppia rappresenta il centroide in ognuno dei due cluster
 - vi sono inseriti (indicati con &) aggettivi simili ad a_1 in un cluster, simili ad a_2 nell'altro

25

AGGETTIVI DESCRITTIVI

Struttura a cluster (esempio)

- **Esempio**
 - { [WET, DRY !], watery &, soggy &, humid &, damp &, moist &, ... }
 - { [DRY, WET !], anhydrous &, arid &, dehydrated &, dried-up &, ... }



26

AGGETTIVI DESCRITTIVI

Aggettivi "graduali"

- Talvolta prendendo due cluster differenti collegati da una relazione di contrarietà troviamo una serie di aggettivi ordinabili secondo una scala di valori
 - **Esempio**
< *ancient*, *OLD*, *middle-aged*, *mature*, *adolescent*, *YOUNG*, *infantile* >
- Dopo uno studio attento, i ricercatori di WordNet hanno deciso di non implementare tali scale in quanto gli aggettivi con tali proprietà rappresentano solo il 2% del totale

27

AGGETTIVI DESCRITTIVI

Colori

- I colori sono casi particolari di aggettivi
 - Possono essere interpretati sia come nomi che come aggettivi
 - Nessun colore ha il proprio contrario
 - Esclusi bianco/nero
 - Non vengono considerati i concetti di complementarietà (verde/rosso)
- In WordNet i colori sono in generale considerati come nomi ed eventualmente accostati all'aggettivo *colored*

28

AGGETTIVI RELAZIONALI

- Sono tutti quegli aggettivi che accostati ad un nome
 - ne cambiano il significato
 - formano una "parola unica" di senso compiuto.
 - ne forniscono una specializzazione
 - Strumento *musicale*
 - Igiene *dentale*
- Non esprimono e non attribuiscono un valore all'oggetto a cui vengono accostati
- Non hanno contrario
- Non possono paragonati in base ad un qualche valore

29

AGGETTIVI RELAZIONALI

Struttura

- La mancanza di contrarietà esclude a priori l'organizzazione a cluster utilizzata per i descrittivi
- Organizzazione insiemistica simile a quella utilizzata per i nomi
- Essendo gli aggettivi relazionali associati a sostantivi in ogni Synset viene inserito un puntatore ai sostantivi a cui può essere associato

30

AGGETTIVI RELAZIONALI

Polisemia

- Alcuni aggettivi relazionali possono assumere sfumature diverse a seconda del nome a cui vengono accostati
 - Criminal law
 - Criminal behavior
- In tutte queste situazioni vengono ovviamente creati due Synset diversi

31

AGGETTIVI

Informazioni aggiuntive

- In generale ci sono aggettivi che, in certi contesti, devono essere posizionati in precisi punti rispetto al sostantivo a cui fanno riferimento
- Tali informazioni sono state aggiunte in WordNet inserendo una lettera al posto di tali aggettivi
- Si distinguono 3 casi:
 - (p): aggettivi che si trovano necessariamente vicino ad un verbo e che fanno quindi la funzione di participio
 - *up, awake, asleep*
 - (a): agg. che si trovano sempre prima del nome
 - *Putative father* e non *father is putative*
 - (ip): agg. che si trovano dopo il sostantivo
 - *President elect*

32

VERBI

Visione generale

- I verbi sono in quantità minore rispetto ai nomi
- Significato molto flessibile
 - cambia a seconda del contesto in cui vengono utilizzati
- Alcuni verbi usati molto frequentemente anche in forma ausiliaria
 - *Be, have, get...*

33

VERBI

Polisemia

- Le forme verbali sono dunque caratterizzati da un'altissima polisemia
- Esempio: *Beat* (colpire)
 - {*beat, strike, hit*} - verbo di contatto
 - {*beat, flatten*} - verbo di cambiamento
 - {*beat, throb, pulse*} - verbo di movimento
 - {*beat, defeat*} - utilizzato in competizioni
 - {*beat, punish*} - comportamenti sociali
 - {*beat, shape*} - nel senso di creare, modellare
 - {*beat, baffle*} - nel senso di impedire
 - {*beat, stir, whisk*} - in ambito culinario
 - {*beat, mark*} - in ambito musicale

34

VERBI

Struttura

- I verbi sono organizzati in maniera del tutto simile a quella utilizzata per i nomi
- Struttura a insiemi di parole tra loro sinonime (Synset)
- Suddivisione in 15 file diversi in base a criteri semantici
- Non c'è una struttura gerarchica simile a quella dei nomi
 - Per ogni verbo c'è un puntatore @-> al file di appartenenza

35

VERBI

Classificazione verbi

- Verbi riguardanti funzioni del corpo
 - dormire, mangiare,...
- Verbi di cambiamento
 - modificare, adattare,...
- Verbi di comunicazione
 - Salutare, insegnare,...
- Verbi di competizione
 - Gareggiare, vincere,...
- Verbi di "consumo"
 - Consumare, rovinare

36

VERBI

Classificazione verbi (2)

- Verbi di contatto
 - Colpire, urtare...
- Verbi cognitivi
 - Imparare, memorizzare...
- Verbi di creazione
 - Inventare, realizzare...
- Verbi di movimento
 - Camminare, volare...
- Verbi emotivi e legati alla psiche
 - Amare, incoraggiare...

37

VERBI

Classificazione verbi (3)

- Verbi di stato
 - Avere, trovarsi...
- Verbi percettivi
 - Annusare, udire...
- Verbi di possesso
 - Prendere, ricevere...
- Verbi di interazione sociale (usati in ambito politico – economico)
 - Educare, promuovere...
- Verbi per descrivere eventi naturali e climatici
 - Piovere, nevicare...

38

VERBI

Induzione e Meronimia

- Il principio di induzione può essere applicato ai verbi?
 - *Russare* implica *dormire*
- Si può pensare ad una sorta di meronimia
 - *Russare* è una componente del *dormire*
- Meronimia basata su aspetti temporali
 - Una azione A può essere suddivisa in una sequenza ordinata A_1, \dots, A_n di azioni più piccole
 - Ogni sottoazione A_k può essere considerata meronima di A.

39

VERBI

Induzione e Troponimia

- Un altro concetto legato all'implicazione è quello di Troponimia
- La Troponimia è il corrispondente dell'Iponimia dei nomi
- Il verbo V_1 è Troponimo di V_2 se V_1 indica un caso particolare del più generico verbo V_2
 - Passeggiare è troponimo di camminare
- La differenza dalla meronimia è che qui V_1 e V_2 avvengono contemporaneamente per tutta la durata dell'azione.

40

VERBI

Altri casi di induzione (1)

- Alcuni verbi implicano altri verbi secondo una considerazione basata su supposizioni temporali (Backward Presupposition)
 - Esempio: i verbi "colpire" e "mancare" implicano, secondo questo modo di pensare, il verbo "mirare" perché si presuppone che l'azione precedente al "colpire" o al "mancare" sia appunto quella di "mirare"

41

VERBI

Altri casi di induzione (2)

- Alcuni verbi implicano ne implicano altri su considerazioni del tipo mittente/ricevente:
 - Esempio: Il verbo "insegnare" implica il verbo "imparare" perché si presuppone che se un soggetto compia l'azione dell'*insegnare* ci sia qualcuno, dall'altra parte ci sia un soggetto che compia quella di "*imparare*"
 - Lo stesso vale per i verbi "parlare"/"ascoltare", "dare"/"ricevere" ecc...
- Ovviamente tutti questi verbi hanno la proprietà fondamentale di essere transitivi

42

VERBI

Implicazioni utilizzate

- Per non complicare comunque la struttura è stato deciso di considerare due tipi di puntatori per i quattro tipi di implicazione
 - ~→ per i rapporti di troponimia e di meronimia
 - *→ per gli altri due casi di implicazione

43

VERBI

Contrarietà

- La relazione di contrarietà risulta un po' difficile da definire nei verbi
- Si assume comunque la definizione comune che un verbo V1 è il contrario di V2 se ne descrive l'azione inversa
 - *Salire / scendere*
 - *Prendere / dare*
- In Inglese, nella maggioranza dei casi, si ottiene applicando dei prefissi quali dis-, un-, de- ecc..
 - *appear / disappear*
- Da notare la differenza con l'implicazione vista nel lucido precedente:
 - Il contrario di un verbo non è necessariamente eseguito da un eventuale oggetto
 - Il rapporto di contrarietà è simmetrico
- Il termine contrario di un verbo viene indicato con !→

44

VERBI

Struttura finale

- Suddivisione in 15 file diversi in base a criteri semantici
- Raggruppamento di verbi in insieme di sinonimi Synset
- Struttura gerarchica difficilmente realizzabile
 - Una tale struttura si spanderebbe molto in ampiezza e poco in profondità
- Realizzazione generale simile a quella degli aggettivi a cluster

45

VERBI

Struttura finale (2)

- Ogni Synset è formato da puntaori per
 - File di appartenenza
 - Implicazioni
 - Troponimie
 - Contrari
 - Eventuali altri verbi correlati per semantica (also see)
- Talvolta possono essere inserite brevi descrizioni e piccole frasi per meglio descrivere la semantica del verbo.

46

AVVERBI

Struttura

- La struttura degli avverbi è relativamente semplice
- Organizzati in Synset senza una propria struttura gerarchica
- Ogni Synset possiede
 - *breve descrizione semantica*
 - *eventuale puntatore all'aggettivo da cui deriva* !→
 - *Eventuale contrario* !→

47

STRUTTURA TECNICA

- L'intero progetto WordNet è suddiviso in quattro parti distinte
 - Realizzazione dei dati da parte dei lessicografi
 - Un software che converte i dati forniti in input per il database di tutti i termini
 - Il database
 - Software per l'accesso e la ricerca nel database
- In generale i programmi software che implementano tale struttura sono scritti in linguaggio C su piattaforma Unix

48

STRUTTURA TECNICA

Grinder utility

- Una sorta di compilatore per i dati forniti dai lessicografi (scritto in C)
 - Analizza i dati dai lessicografi
 - Traduce i dati ottenuti in una sintassi accettata dal DataBase di WordNet
 - Costruisce i puntatori interni di tali dati

49

STRUTTURA TECNICA

Database di WordNet

- Per ogni categoria sintattica abbiamo due file
 - Index.pos
 - Data.pos
- Scritti in formato ASCII e quindi molto comprensibili anche da una persona umana
- Liberamente disponibili
- Il primo file è un indice di tutti i Synset presenti della categoria sintattica e presenti in Data.pos

50

WORDNET E SVILUPPI FUTURI

- WordNet è disponibile sia in versione on-line che liberamente scaricabile all'indirizzo: <http://wordnet.princeton.edu/>
- Il successo di WordNet ha spinto molti ricercatori ad estendere lo sviluppo di un tale supporto anche per lingue differenti da quella inglese (MultiWordNet)
<http://multiwordnet.itc.it/english/home.php>
- Tra gli sviluppi futuri c'è anche la realizzazione di un Italian WordNet
<http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=834/vers=ita>

51

BIBLIOGRAFIA

- **Introduction to WordNet: An On-Line Lexical Database** – *George A. Miller, Christiane Fellbaum*
- **WordNet an electronic lexical database** – *Christiane Fellbaum* – The Mit Press
- <http://wordnet.princeton.edu/>
- <http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=834/vers=ita>