Le lingue al crocevia di dati, informazioni e conoscenza

Roberto Basili

(Università di Roma, Tor Vergata, basili@info.uniroma2.it)

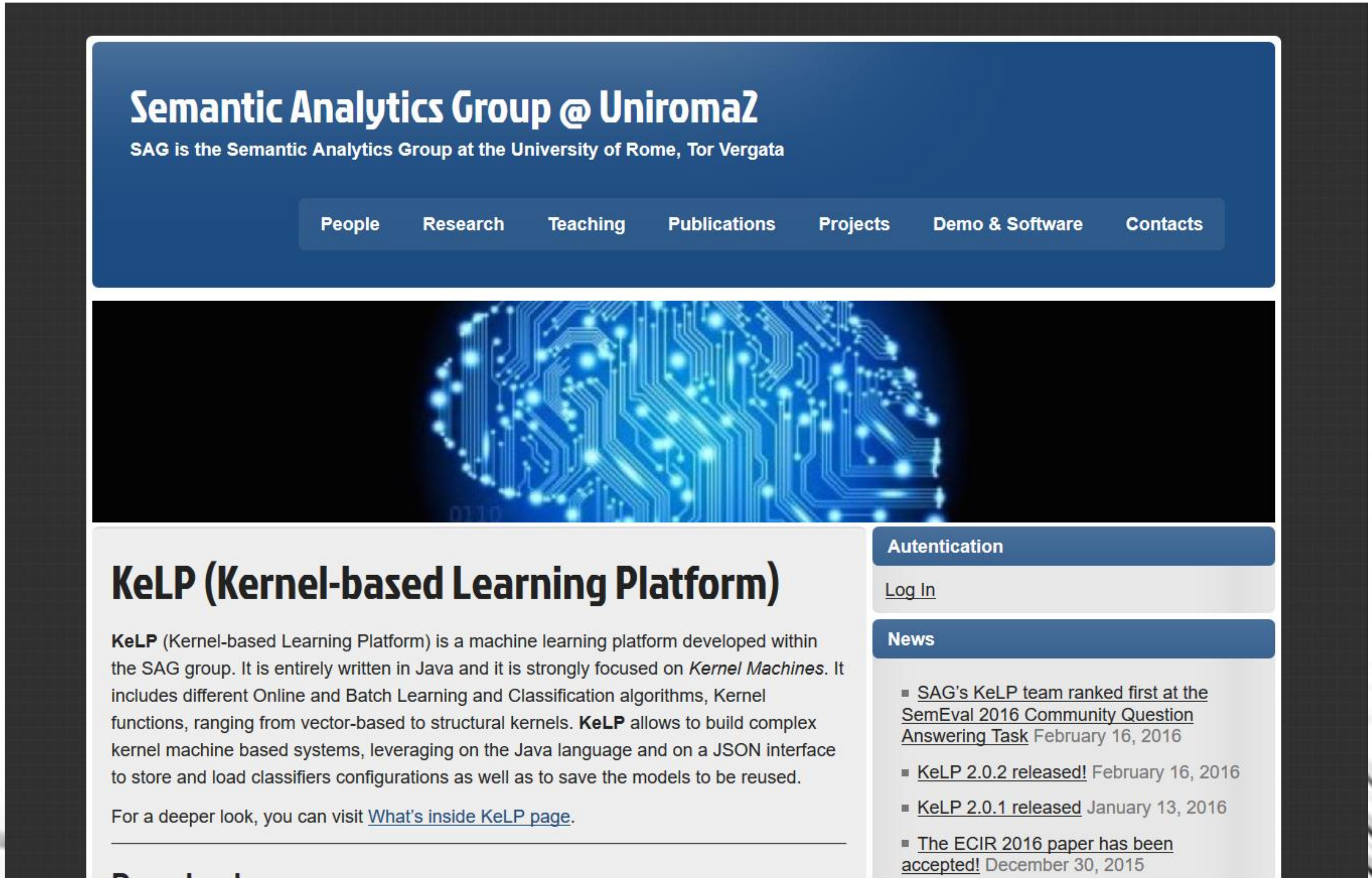
dblp: http://dblp.uni-trier.de/pers/hd/b/Basili:Roberto.html

Google scholar: https://scholar.google.com/citations?user=U1A22fYAAAAJ&hl=it&oi=sra

Bologna, 17 Maggio 2018

Semantic Analytics Group

- http://sag.art.uniroma2.it



Semantic Analytics Group @ Uniroma2

SAG is the Semantic Analytics Group at the University of Rome, Tor Vergata

People | Research | Teaching | Publications | Projects | Demo & Software | Contacts

- Content Processing an Acquisition
- Ontology Engineering
- Machine Learning
- Web & Information Retrieval

- Text Processing and Natural Language Parsing
- Distributional Semantics
- Human-Robot Interaction
- Semantic Role Labeling
- Sentiment Analysis

People

Professors | Postdocs | PhD Students | Fellows

Paolo Annesi

Postdoc

annesi@info.uniroma2.it

Details...

Autentication

Log In

News

- SAG's KeLP team ranked first at the SemEval 2016 Community Question Answering Task February 16, 2016
- KeLP 2.0.2 released! February 16, 2016
- KeLP 2.0.1 released January 13, 2016
- The ECIR 2016 paper has been accepted! December 30, 2015

SAG in ACL Benchmarking Champaigns

- SEMEVAL 2016 (and 2017): Community Question Answering (English) Best system
- Sentipolc 2014: Sentiment Analysis on Twitter (Ita) (Best system on Irony Detection)
- SemEval 2014: Aspect-based SA (Eng)
- Best paper COLING 2014
- Best Innovative System (Robocup, 2014)
- SemEval 2013:
  - Task: Spatial Role Labeling (Eng/Robotics)
  - Task: Structured Text Similarity (Eng)
  - Task: Opinion Analysis over Twitter (Eng)
- StarSem Shared Task 2012: Text Similarity
- EvalIta: Frame Labeling over Italian Texts (Ita)

Outline

- **Artificial Intelligence & Natural Language Processing**
  - Comunicazione linguistica & Conoscenza
  - Il ruolo dei dati
- **Natural Language Processing: *Task*, Modelli e Metodi**
- **Un esempio: computational semantics in Prolog**
- **Trattamento delle lingue e *Machine Learning***
  - Statistical Language Processing
  - Apprendimento discriminativo per l'NLP
- **Natural Language Processing: applications**
- **Conclusions & Perspectives**

AI (*aka* IA): un fenomeno sociale

Lo sviluppo continuo del campo dell'Intelligenza Artificiale sposta in avanti la frontiera della cosiddetta **machine intelligence**.

Ironicamente si rafforza l'**"effetto IA"** (**"odd paradox"**)

- Non appena l'IA porta nuovi risultati nella vita comune, le persone si abituano a tali tecnologie, **e smettono di considerarle IA**. Questo schema si ripete.
- IA non rilascia prodotti dirompenti dal nulla. Piuttosto, le tecnologie IA costruiscono in modo incrementale approssimazioni sempre migliori dell'intelligenza

Nonostante questo, l'intelligenza umana è la scelta d'elezione per la ***misura del progresso dell'IA***.

Tor Vergata

AI: the pendulum

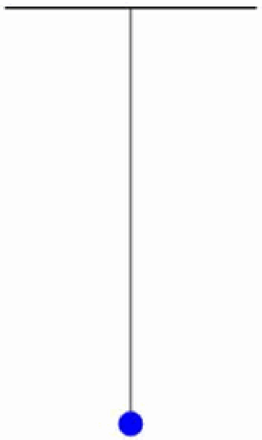
- "A physical symbol system has the necessary and sufficient means for general intelligent action.

  --Allen Newell &

  Herbert Simon

- Symbols are Luminiferous Aether of AI

  —Geoff Hinton

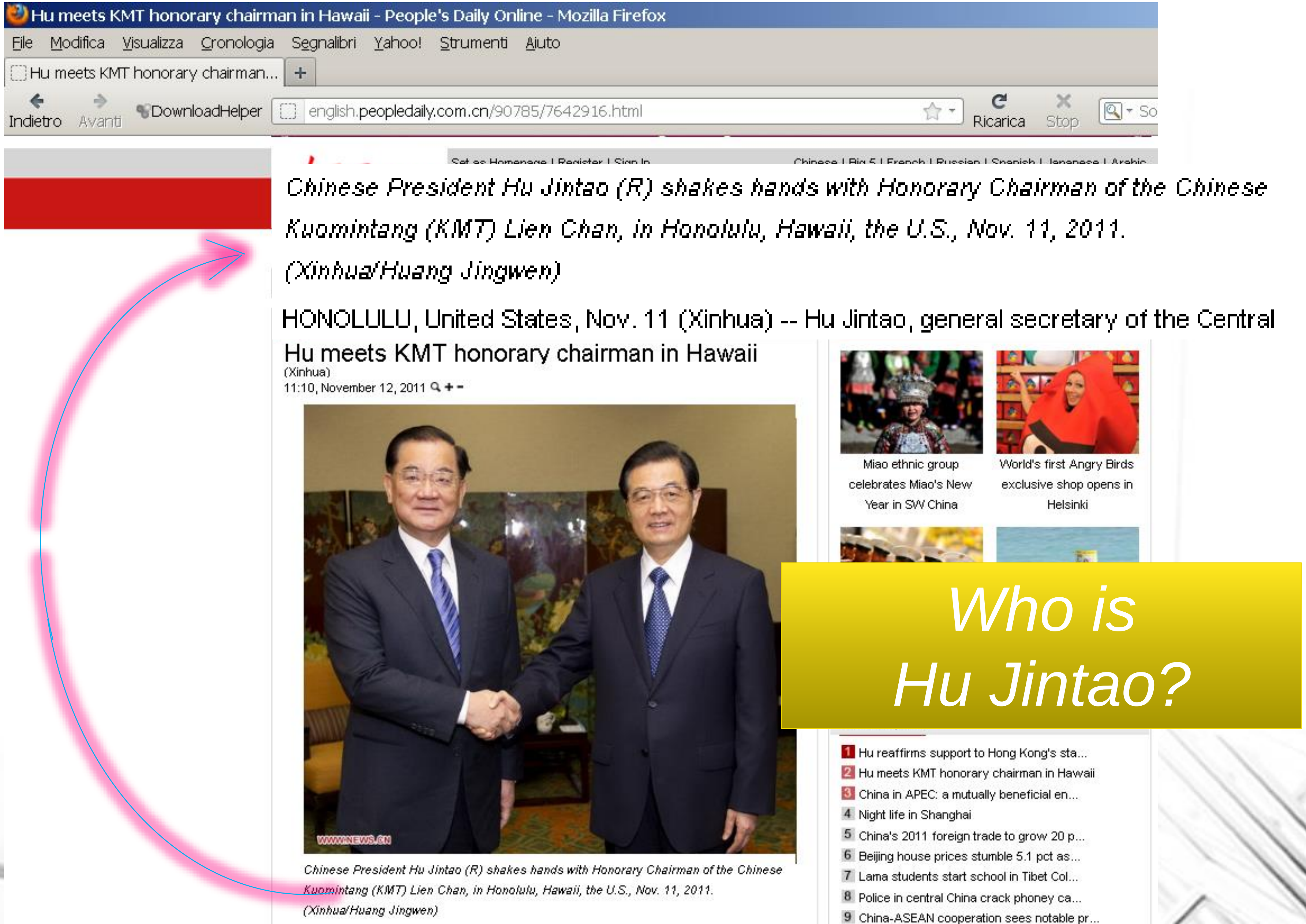


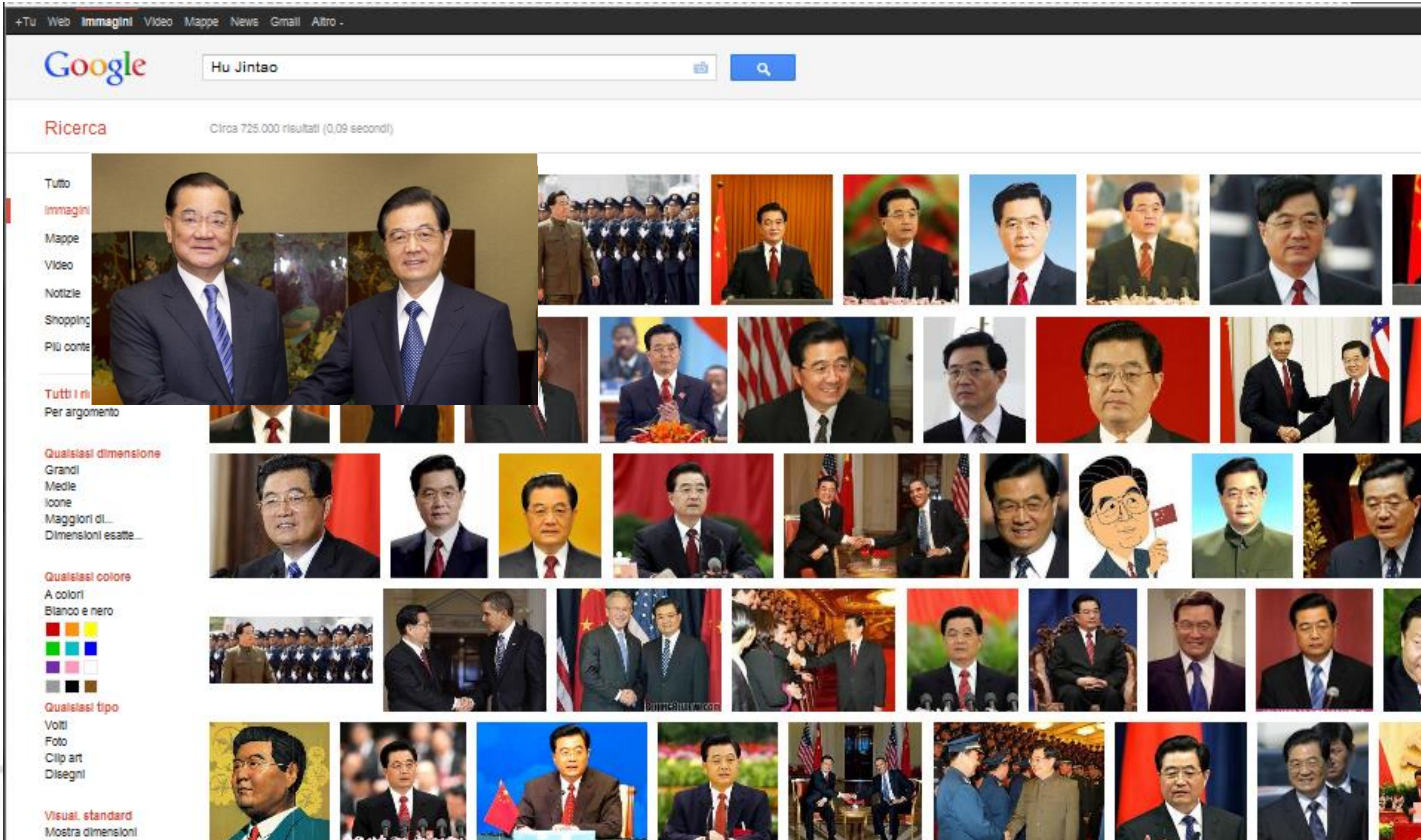
Tor Vergata

Semantics, Open Data & Natural Language

Digital contents are mostly ***opaque*** from a **semantic standpoint**

Information, Web and Natural Languages




+Tu Web Immagini Video Mappe News Gmail Altro -
Google
Hu Jintao
Ricerca
Circa 725.000 risultati (0,09 secondi)
Tutto
Immagini
Mappe
Video
Notizie
Per argomento
Qualsiasi dimensione
Grandi
Medie
Icone
Maggiori di...
Dimensioni esatte...
Qualsiasi colore
A colori
Bianco e nero
Qualsiasi tipo
Volti
Foto
Clip art
Disegni
Visual. standard
Mostra dimensioni

Content Semantics & Natural Language

- Le lingue sono i veicoli tradizionali e consolidati della creazione, condivisione e comunicazione delle conoscenze diffuse nei contenuti del Web

- Sono le parole e le strutture sintattiche ad esprimere in modo trasparente i concetti, le attività, gli eventi, le astrazioni e le relazioni concettuali che noi scambiamo attraverso i flussi di dati

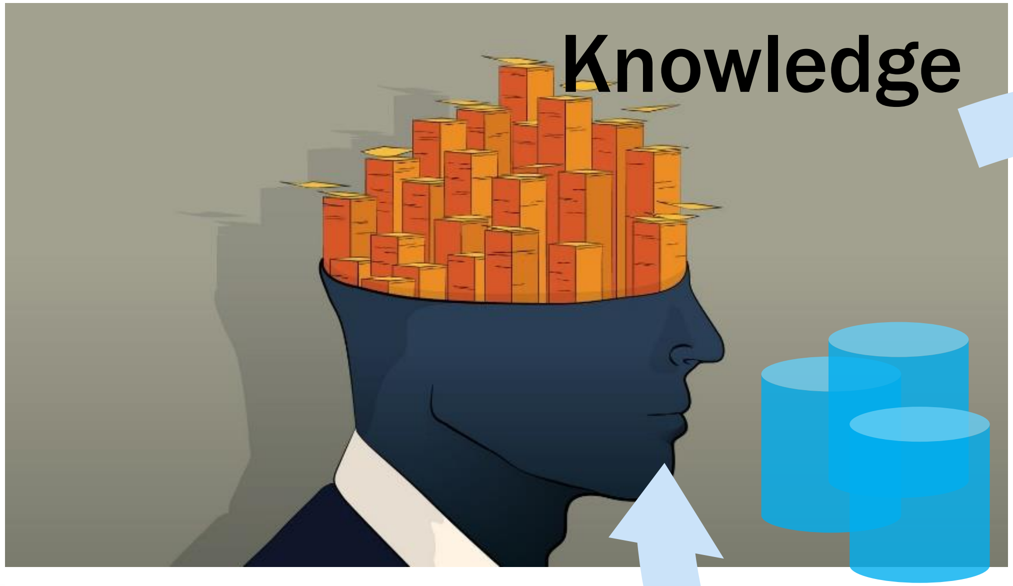
- "*Language is parasitic to knowledge representation languages but the viceversa is not true*" (Wilks, 2001)

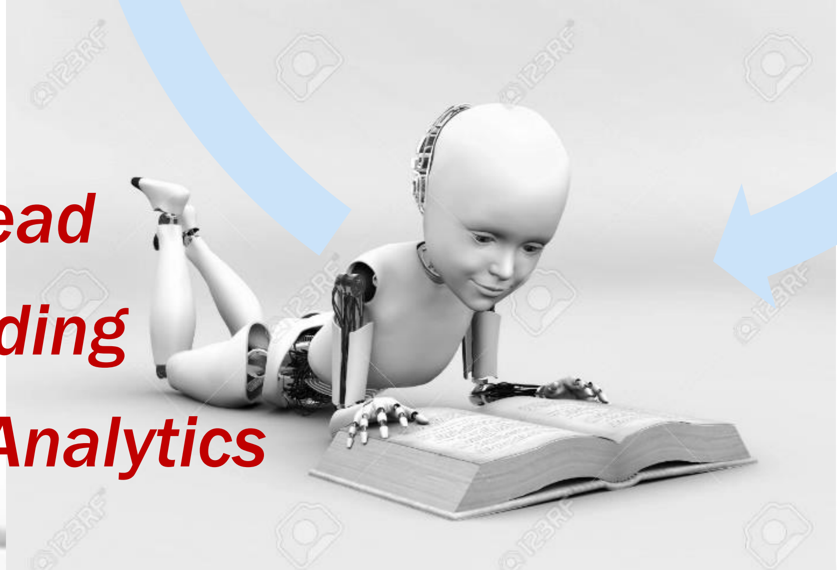
- Apprendere la lettura (**Learning to Read)** abilita la genesi di nuova conoscenza (**Knowledge Distillation) attraverso processi integrati di Interpretazione semantica**

Semantica, Lingue & Learning: tasks

- Dal **Learning to Read** alla **Knowledge Distillation:**
  - **Information Extraction**
    - Entity Recognition and Classification
    - Relation Extraction
    - Semantic Role Labeling (Shallow Semantic Parsing)
  - **Estimation of Text Similarity**
    - Structured Text Similarity/Textual Entailment Recognition
    - Sense disambiguation
  - **Semantic Search, Question Classification and Answer Ranking**
  - **Knowledge Acquisition**, e.g. ontology learning
  - **Social Network Analysis, Opinion Mining**

AI & NLP: knowledge acquisition, and decision-making





Learning to read

Machine Reading

Learning for Analytics

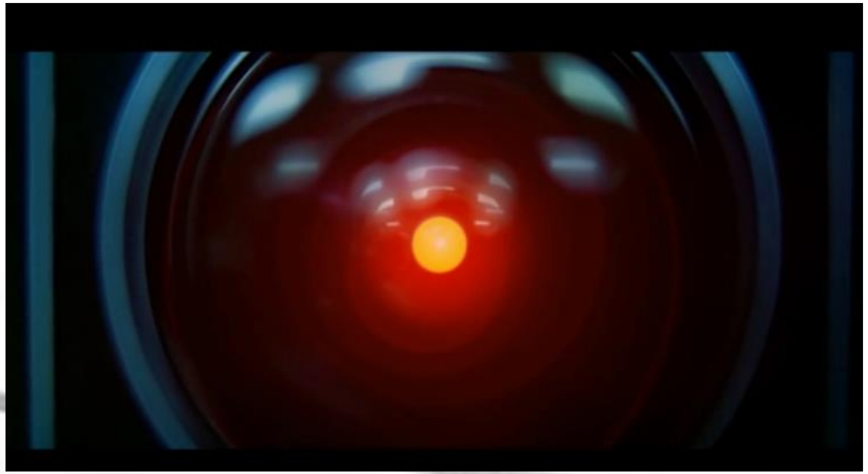
Outline

- Artificial Intelligence & Natural Language Processing
  - Comunicazione linguistica & Conoscenza
  - Il ruolo dei dati
- Natural Language Processing: *Task*, Modelli e Metodi
- Trattamento delle lingue e *Machine Learning*
  - Statistical Language Processing
  - Apprendimento discriminativo per l’NLP
- *Natural Language Processing*: applicazioni
- Conclusioni & Prospettive

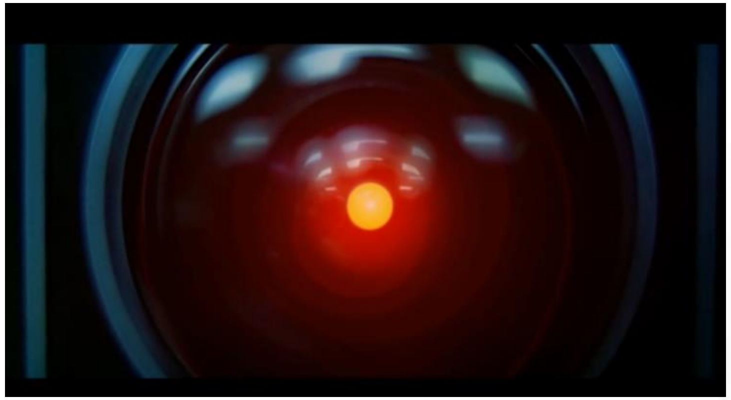
NLP: quali conoscenze?

- HAL 9000, da “*2001: A Space Odyssey*”

- Dave: *Open the pod bay doors, Hal*.
- HAL: *I’m sorry Dave, I’m afraid I can’t do that*.



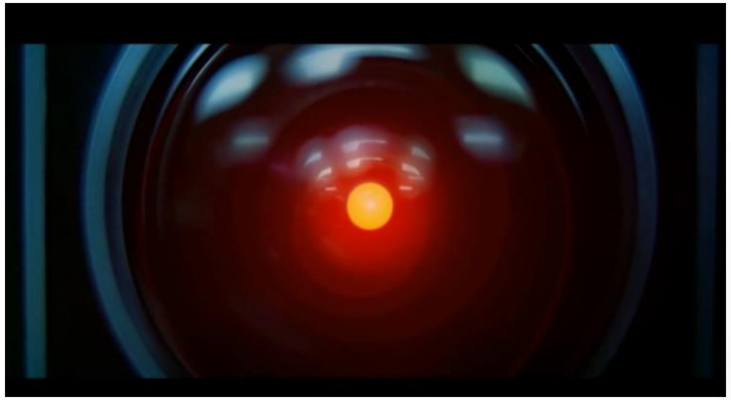
Qual’e’ la conoscenza di HAL?



- **Riconoscimento e Sintesi del linguaggio parlato**
  - Dizionari (*spelling*)
  - Regole fonetiche (come i suoni vengono riconosciuti e prodotti)

- **Comprensione**
  - **Conoscenze Lessicali**
    - Qual’e’ il significato delle parole?
    - Come tali significati si combinano (`*pod bay door*’)
  - **Competenza riguardo alla struttura sintagmatica delle frasi**
    - *I’m I do, Sorry that afraid Dave I’m can’t*

Qual’e’ la conoscenza di HAL?



- **Dialogo & pragmatica**
  - “*open the door*” è una richiesta (e non una dichiarazione o una interrogazione)
  - La replica implica una azione ed è necessario usare modi gentili (anche a fronte dell’intenzione di uccidere …)
  - E’ utile comportarsi in modo cooperativo (*I’m afraid, I can’t…*)
  - Infine: cosa significa `*that*’ in `*I can’t do that*’?

Trattamento delle lingue come processo di interpretazione (semantica)

- Elaborare un testo corrisponde a comprendere diversi aspetti relativi al suo significato:
  - Dominio tematico (e.g. scienze/economia/sport)
  - Obbiettivi Operativi (e.g. e-mail spam)
  - Entità coinvolte, ad esempio *persone* or *luoghi*
  - Eventi potenziali (e.g. fatti raccontati dal testo)
  - Obbiettivi Comunicativi (e.g. dialogo, ordini/dichiarazioni/pianificazione)

- RISULTATO: una *rappresentazione esplicita del significato del testo* con lo scopo di *sostenere tipi diversi di decisioni* (inferenze)
  (e.g. ranking nell'IR, pianificazione, acquisizione di nuova conoscenza, …)

Sfide Principali

- Accuratezza Linguistica (i.e. grado di approssimazione della performance dei parlanti nativi)
- Robustezza (errori/rumore/incompletezza)
- Scala
  - Copertura dei fenomeni (Lessici/Grammatiche)
- Expressività
  - Informazione semantica nei Dizionari, Lessici e nei Thesauri
  - Modelli del Mondo e tipi di inferenza
- Flessibilità
  - Variabilità linguistica (e.g. producer vs. consumer)
- Naturalezza
  - Accuratezza percepita

Lingue & Ambiguità

Ambiguità

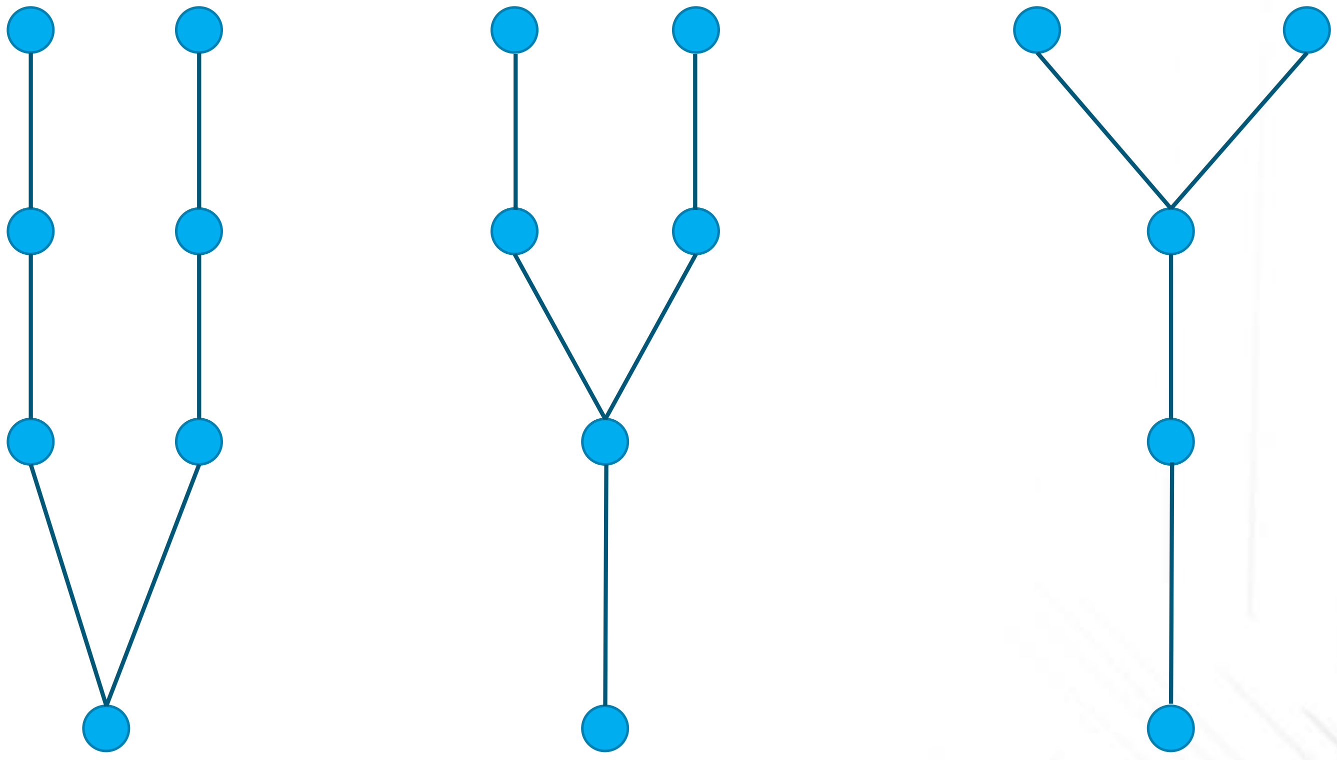
"Dogs must be carried on this escalator"

can be interpreted in a number of ways:

- All dogs should have a chance to go on this wonderful escalator ride
- This escalator is for dog-holders only
- You can't carry your pet on the other escalators
- When riding with a pet, carry it

Livelli di Ambiguità

- Semantica
- Grammaticale
- Morfologica
- Fonologica

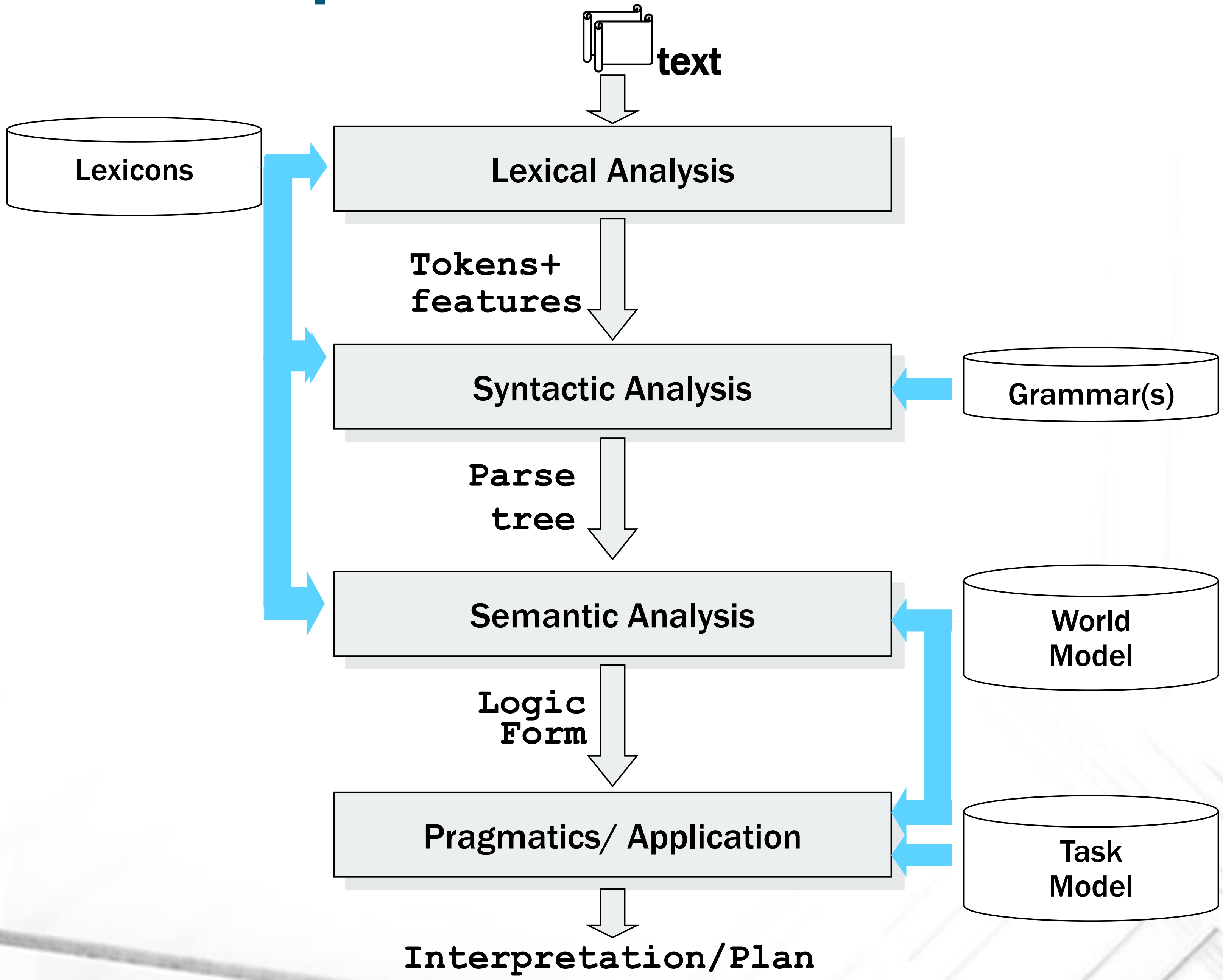


dei/dèi

compro la borsa
in pelle

il timore dei manager

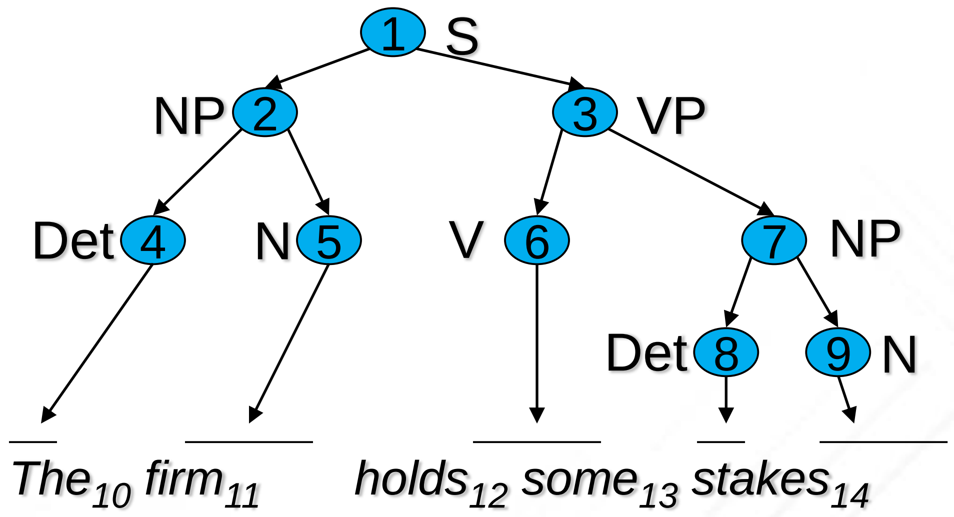
NLP: il processo



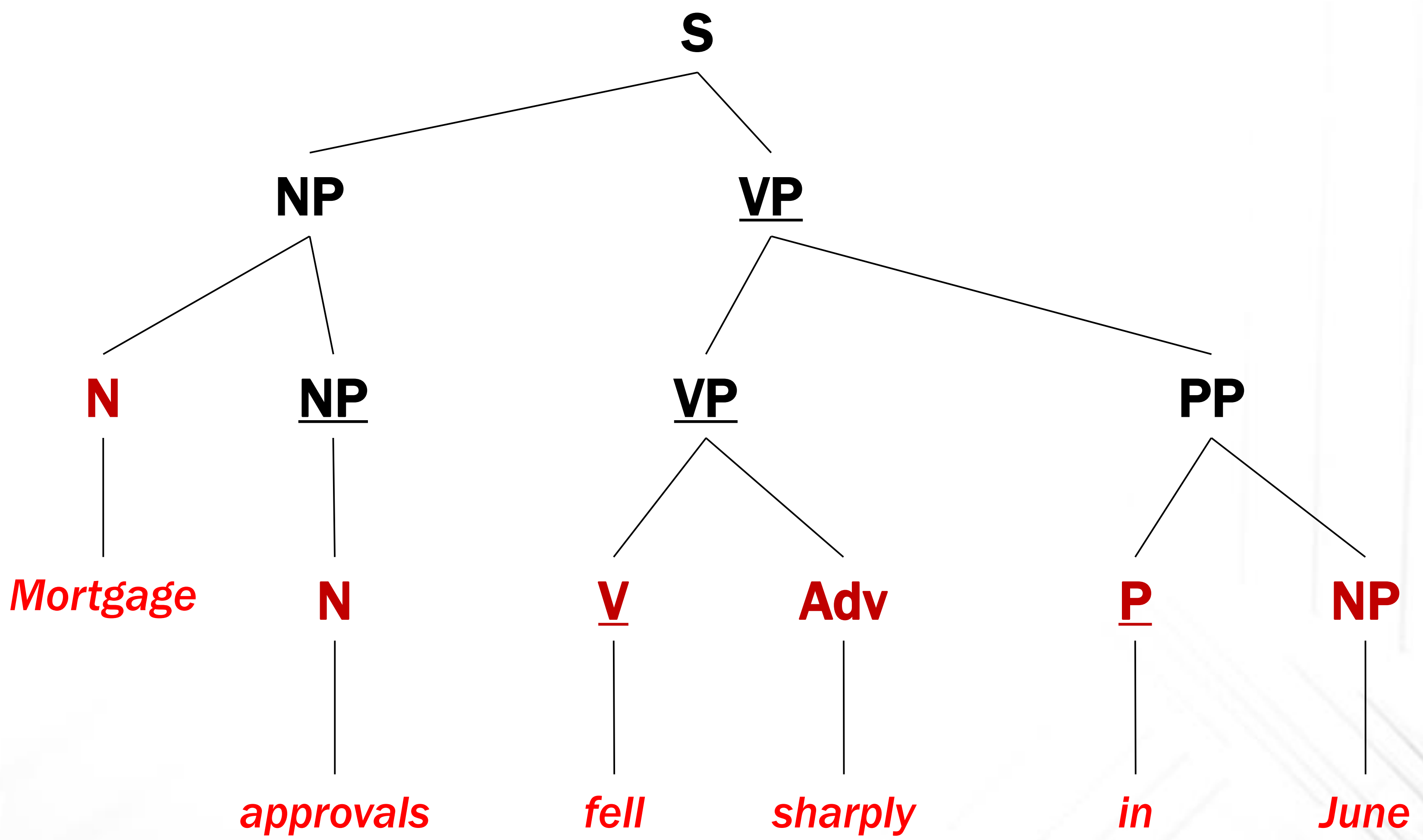
Sintassi: Grammatiche a struttura sintagmatica

"*The firm holds some stakes*"

- **Vn={S,NP,VP,Det,N}, Axiom: S**
- **Productions: {S→NP VP, VP→V NP, NP→Det N}**
- **Derivation:**
  - S > NP VP > Det N VP > *The* N VP > *The firm* VP > *The firm* V NP > *The firm holds* NP > *The firm holds* Det N > *The firm holds some* N > *The firm holds some stakes*



Parsing a Costituenti (con Head segnalate)



Altre strutture: Dependency Parsing

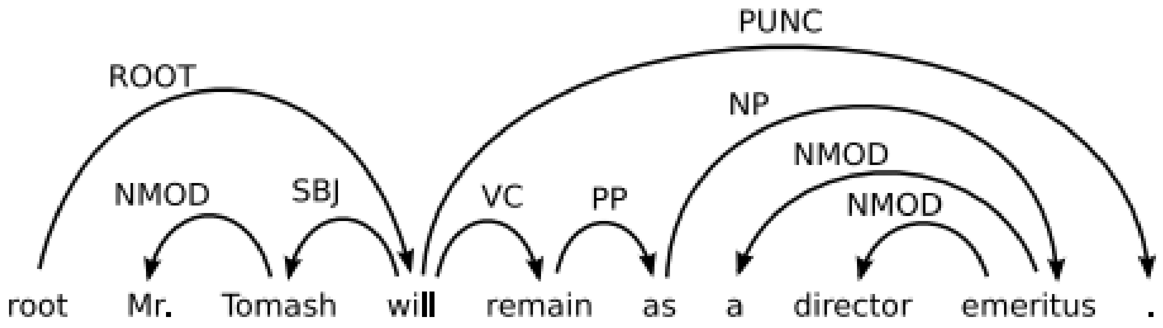


Figure 1: A projective dependency graph.

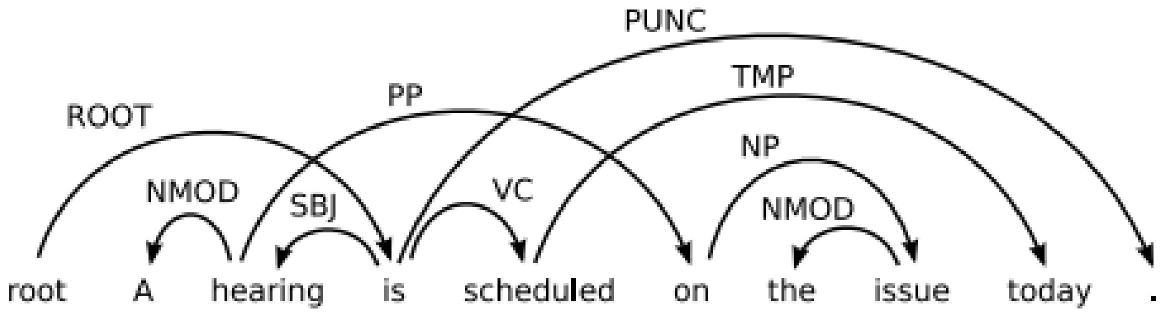
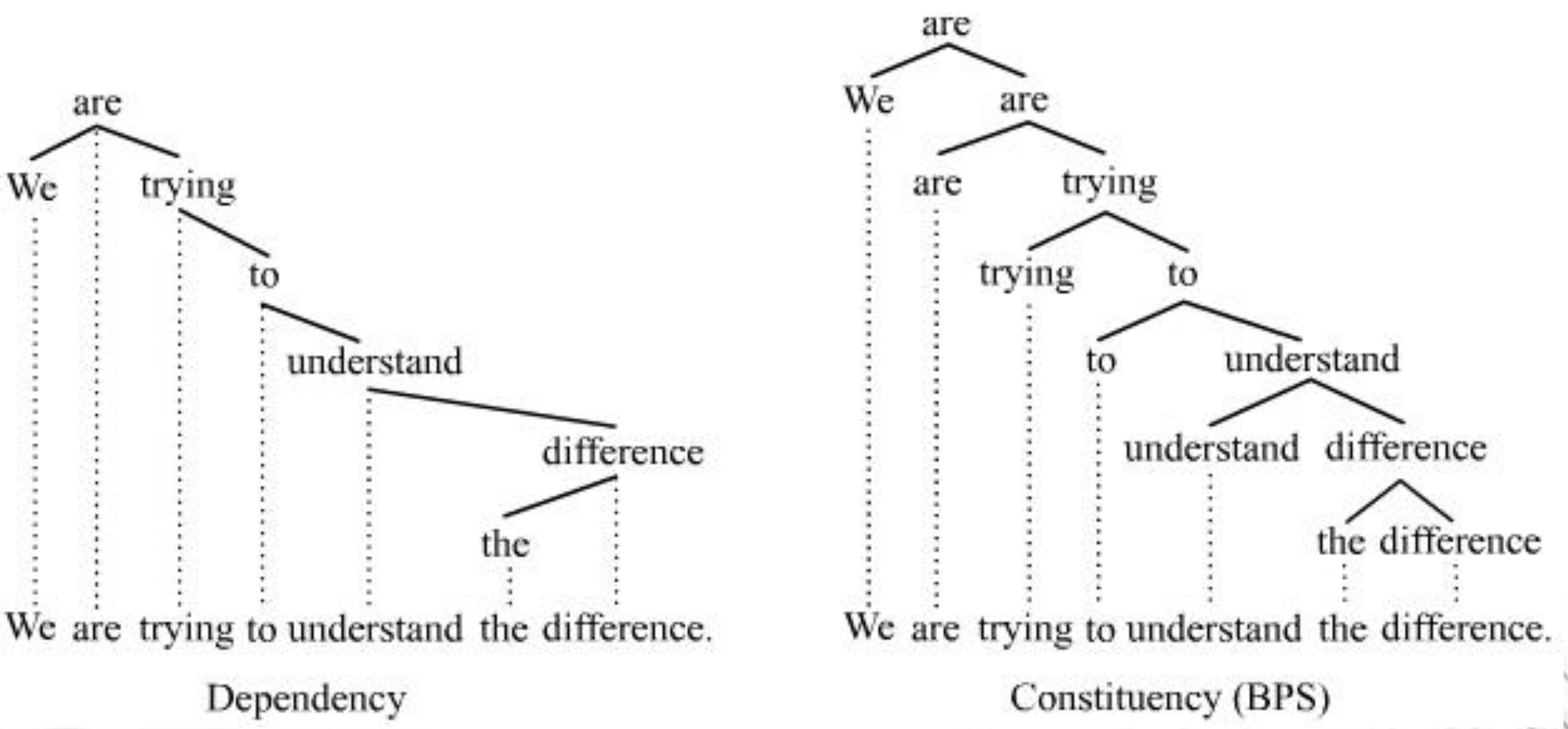


Figure 2: Non-projective dependency graph.

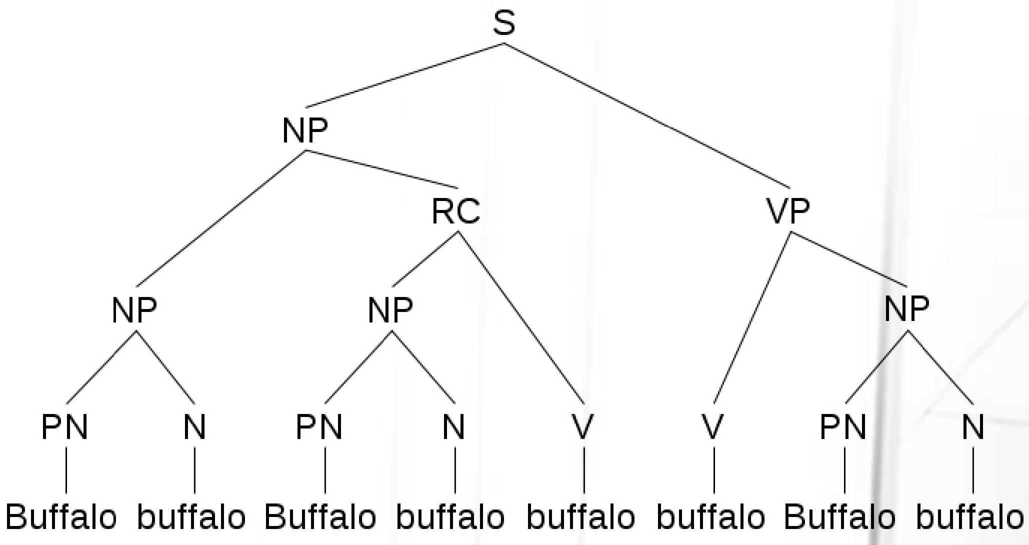
Constituency vs. Dependency



Parsing e Ambiguità

- Lo spazio di ricerca di un parser è enorme a causa delle molteplici ambiguità che interagiscono in modo combinatorio
  - E.g. *La vecchia porta la sbarra*,

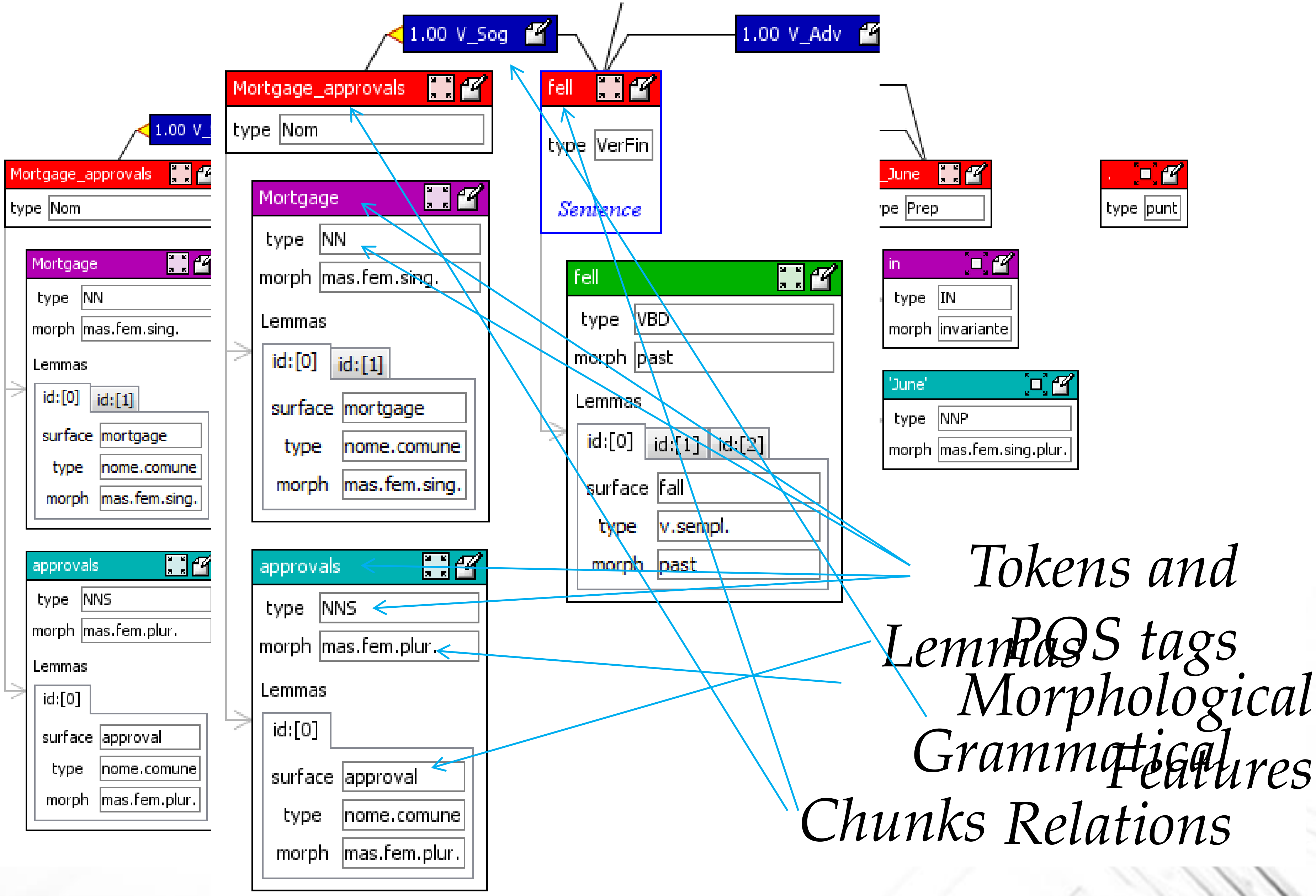
Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo



Bison from Buffalo, New York who are intimidated by other bison in their community also happen to intimidate other bison in their community

- C'e una dipendenza stretta con la semantica
  - La maggior parte di ambiguità non si può risolvere a livello grammaticale
  - L'informazione lessicali (i sensi) sono cruciali
    - *Operare in un mercato* $\neq$ *Operare un paziente*


(A((SHIP SHIPPING)SHIP)
SHIPPING(SHIPPING SHIPS))



FT (July, 29): *Mortgage approvals fell sharply in June.*

Semantica

- Qual è il significato di una frase come

  John saw Kim

- Proprietà desiderabili di una rappresentazione del significato:
  - **Deve essere composizionale,** i.e. il significato deve essere funzione dei costituenti come *Kim*, *John* e il verbo *see*
  - **Independente dai fenomeni della sintassi**, e.g. *Kim was seen by John* è una frase sinonima (parafrasi)
  - Deve essere usata per **derivare delle conseguenze**:
    - *Who was seen by John? Kim!*
    - *John saw Kim. He started running to her.*

Logica e Semantica delle Lingue

John saw Kim

S

saw(s,k)

NP

s

Sam

s

VP

$\{ x : saw(x,k)\}$

V

$\{ \langle x,y \rangle : saw(x,y)\}$

saw

$\{ \langle x,y \rangle : saw(x,y)\}$

NP

k

Kim

k

Computational Semantics

- See slides on «*Compositional Semantics in Prolog*»

Prospettive Linguistiche sul Significato

- Semantica Lessicale
  - Studio del significato delle parole individuali
- Semantica Formale (o Compositional Semantics, Sentential Semantics)
  - Come i significati individuali si compongono al fine di determinare il significato degli enunciate individuali
- Discorso o Pragmatica
  - Come i significati frasali si combinano tra loro e con altri fatti riguardo a diversi tipi di contesti in modo da comporre significati di un testo o di un discorso
  - Il Dialogo o la Conversazione spesso sono integrati nella interpretazione di un Discorso

Lexical Semantics: Synonymy

- Parole che hanno lo stesso significato in alcuni o tutti I contesti
  - *couch / sofa*
  - *big / large*
  - *automobile / car*
  - *vomit / throw up*
  - *Water / H_20*
- Due lessemi sono sinonimi se possono essere sostituiti l’uno all’altro in tutte le situazioni
  - Hanno in questo modo lo stesso **significato proposizionale**

Lexical Semantics: Synonymy (2)

- Gli esempi di sinonimia perfetta sono pochi (o inesistenti)
  - Perché è così?
  - Anche se gli aspetti del significato si preservano potrebbero differire nella accettabilità legata al genere, allo slang, al tono o allo stile

- Example:
  - Water vs. H_2O
  - Non diremmo mai:

    I like fresh H_2O after the tennis

Terminologia

- Lemmi e parole
  - Un **lessema** è la relazione stabile (astratta) tra significato e forma
  - Un **lemma** è la forma grammaticale usata per rappresentare un **lessema.**
    - ***Carpet*** è il lemma di ***carpets, Dormir*** è il lemma di ***duermes*.**
  - Le forme superficiali quali *carpets, duermes* sono le **parole flesse** (**wordforms**)

- Il lemma *bank* ha due **sensi:**
  - Instead, a **bank** can hold the investments in a custodial account in the client's name
  - But as agriculture burgeons on the east **bank**, the river will shrink even more.
- Un **senso** è la rappresentazione discreta di un aspetto del significato di una parola

Sinonimia come relazione tra i sensi piuttosto che tra *parole*

- Ad esempio: *big* and *large*
- Sono sinonimi?
  - *How **big** is that plane?*
  - *Would I be flying on a **large** or small plane?*
- E qui?
  - Miss Nelson, for instance, became a kind of **big** sister to Benjamin.
  - ?Miss Nelson, for instance, became a kind of **large** sister to Benjamin.
- Infatti:
  - *big* ha un senso che significa *being older*, or *grown up*
  - *large* manca del tutto di un tale senso

WordNet

- Home page:
  - http://wordnetweb.princeton.edu/perl/webwn

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: meaning Search WordNet

Display Options: (Select option to change) Change

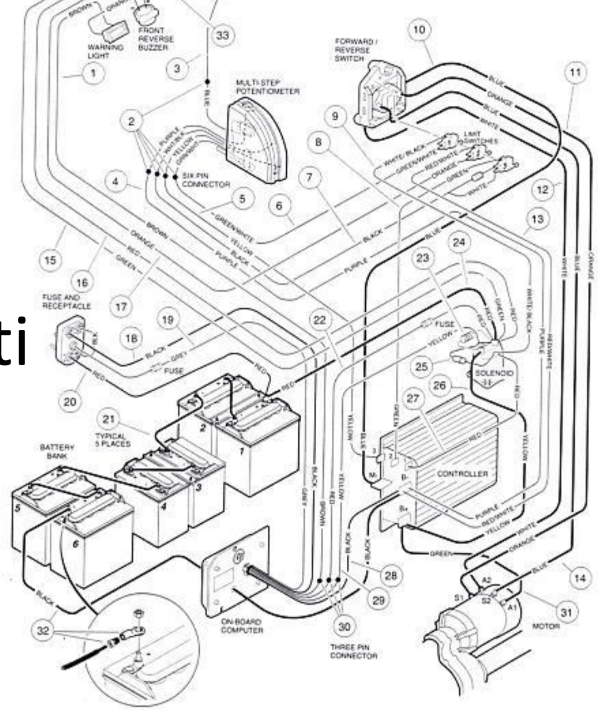
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

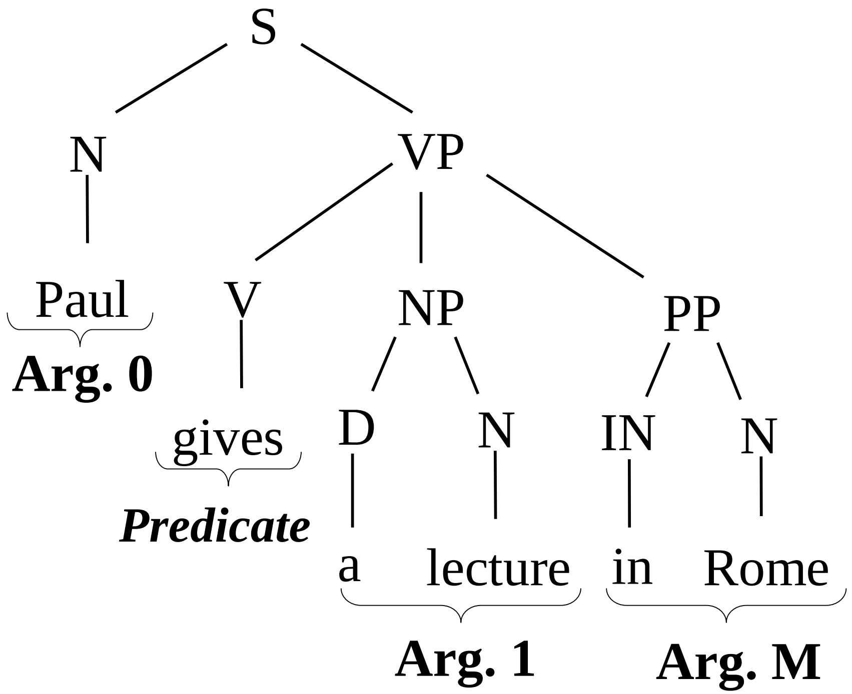
Noun

- S: (n) **meaning**, significance, signification, import (the message that is intended or expressed or signified) *"what is the meaning of this sentence"; "the significance of a red traffic light"; "the signification of Chinese characters"; "the import of his announcement was ambiguous"*
  - *direct hyponym* / *full hyponym*
  - *direct hypernym* / *inherited hypernym* / *sister term*
  - *derivationally related form*
- S: (n) **meaning**, substance (the idea that is intended) *"What is the meaning of this proverb?"*

Dalla semantic lessicale alla logica formale: *compositionality*



- Il *mapping* sintassi-semantica: assemblare significati

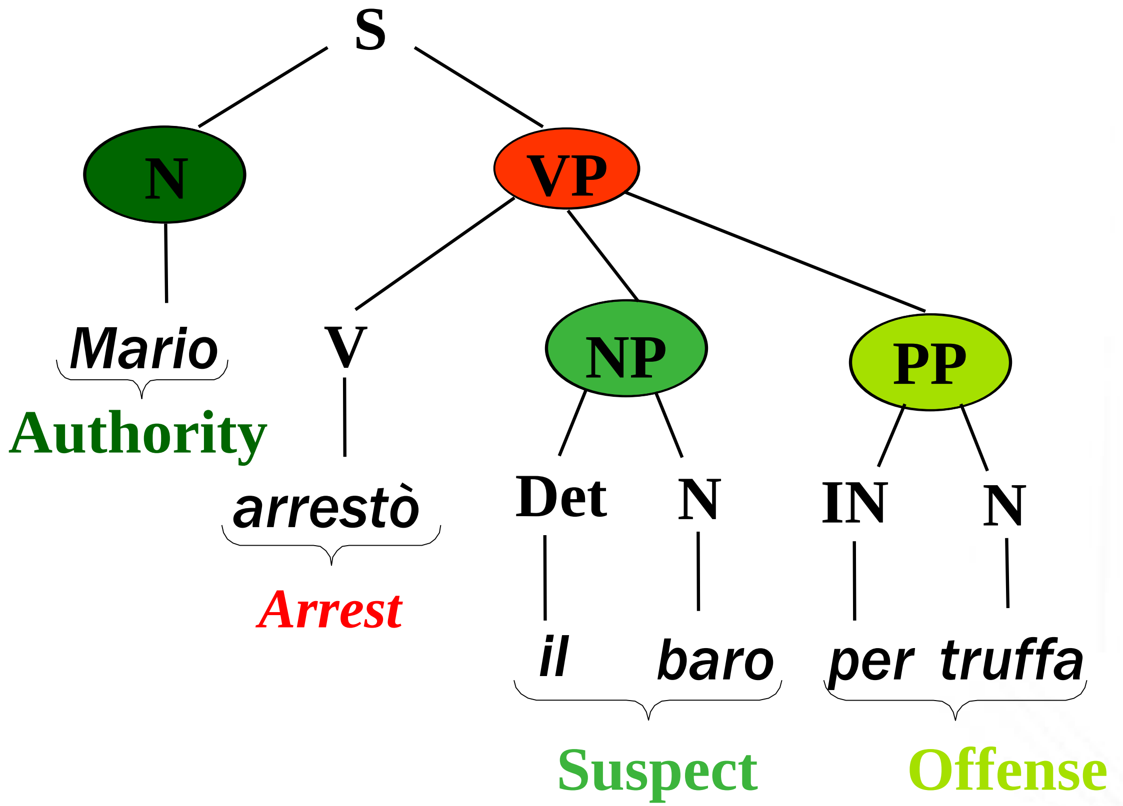


E' richiesta una teoria della predicazione e dei ruoli semantici di riferimento (Teorie *lessicali*: PropBank, Teorie Cognitive: ***Frame***Net)

L'interfaccia sintassi-semantica: Frame Semantics

(Fillmore, 1975)

Mario arrestò il baro per truffa



[Il baro]$_{\text{Suspect}}$ *[fu arrestato]*$_{\text{Arrest}}$ *[da Mario]*$_{\text{Authority}}$ *[per truffa]*$_{\text{Offense}}$

Frame Semantics

- Research in Empirical Semantics suggests that words represents categories of experience (*situations*)
- A *frame* is a cognitive structuring device (i.e. a kind of prototype) indexed by *words* and used to support understanding (Fillmore, 1975)
  - *Lexical Units* evoke a Frame in a sentence
- Frames are made of *elements* that express participants to the situation (*Frame Elements*)
- During communication LUs evoke the frames

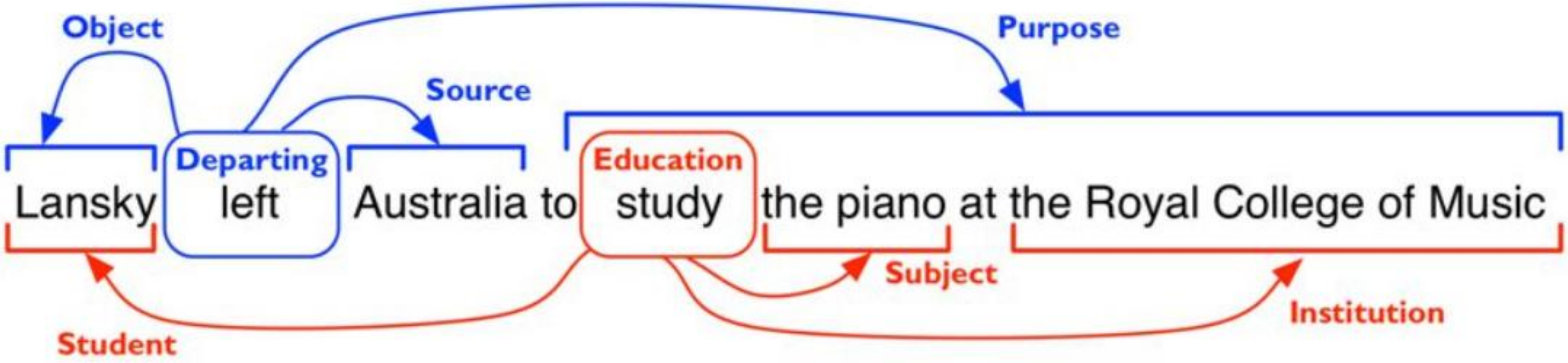
Fram

Frame: KILLING		
A KILLER or CAUSE causes the death of the VICTIM.		
Frame Elements	KILLER	**John** drowned Martha.
	VICTIM	John drowned **Martha**.
	MEANS	The flood exterminated the rats **by cutting off access to food**.
	CAUSE	**The rockslide** killed nearly half of the climbers.
	INSTRUMENT	It's difficult to suicide **with only a pocketknife**.
Predicates	annihilate.v, annihilation.n, asphyxiate.v,assassin.n, assassinate.v, assassination.n, behead.v, beheading.n, blood-bath.n, butcher.v, butchery.n, carnage.n, crucifixion.n, crucify.v, deadly.a, decapitate.v, decapitation.n, destroy.v, dispatch.v, drown.v, eliminate.v, euthanasia.n, euthanize.v, . . .	

Frame Semantics

- Lexical descriptions are expected to define the indexed frame and the frame elements with their realization at the syntactic level:
  - *John bought a computer from Janice for 1000 $*
- Mapping into syntactic arguments
  - the buyer is (usually) in the subject position
- Obligatory vs. optional arguments
- Selectional preferences
  - The seller and the buyer are usually “humans” or “social groups”

Frames: Nuove strutture

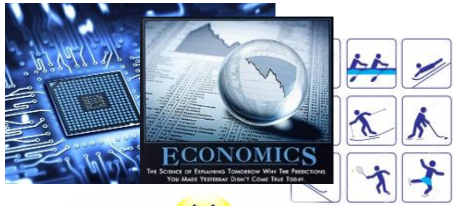


Applicazioni: Fenomeni Semantici di interesse

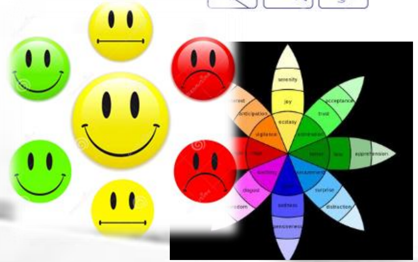
- **Entità.** Individui, luoghi, organizzazioni citate nei testi

- **Relazioni.** Associazioni tra entità

- **Fatti.** Fenomeni o Eventi



- **Topic.** Argomenti di Discussione / Community / Niche Domains



- **Tratti Emotivi e Psicologici.** Social Science, Profilazione

Outline

- **Artificial Intelligence & Natural Language Processing**
  - Comunicazione linguistica & Conoscenza
  - Il ruolo dei dati
- **Natural Language Processing: *Task*, Modelli e Metodi**
- **Esempio: Computational Semantics in Prolog**
- **Trattamento delle lingue e *Machine Learning***
  - Statistical Language Processing
  - Apprendimento discriminativo per l'NLP
- **Natural Language Processing: applications**
- **Conclusions & Perspectives**

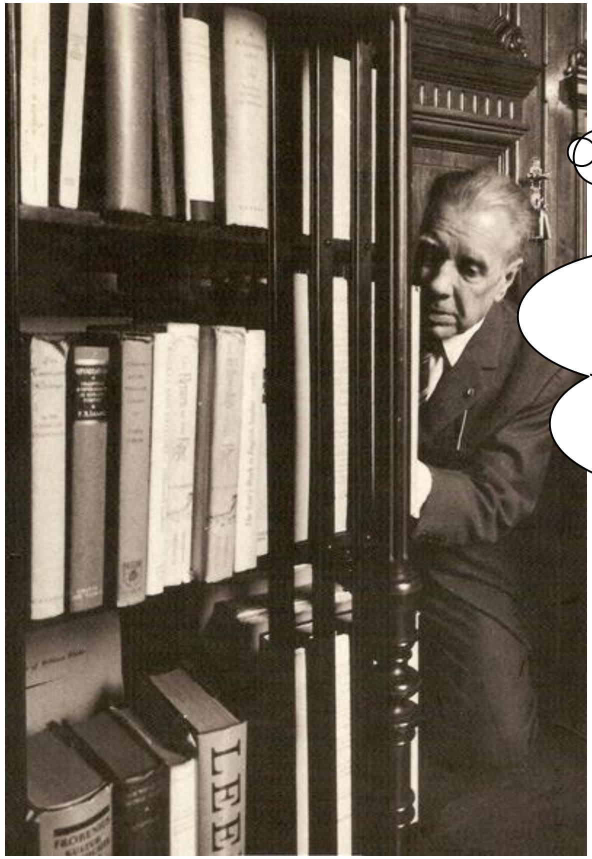
Semantica & Calcolo

- Vedi slide su «Computational Semantics in Prolog»

Outline

- **Artificial Intelligence & Natural Language Processing**
  - Comunicazione linguistica & Conoscenza
  - Il ruolo dei dati
- **Natural Language Processing: *Task*, Modelli e Metodi**
- **Esempio: Computational Semantics in Prolog**
- **Trattamento delle lingue e *Machine Learning***
  - Statistical Language Processing
  - Apprendimento discriminativo per l'NLP
- **Natural Language Processing: applications**
- **Conclusions & Perspectives**

Le lingue tra codici ed esperienza



*… comincia qui la mia disperazione di scrittore. Ogni **linguaggio** è **un alfabeto di simboli** il cui uso presuppone **un passato che gli interlocutori condividono; come trasmettere agli altri l'infinito Aleph** che la mia timorosa memoria a stento abbraccia?*

(*) J.L.Borges, "*L'aleph*", 1949.

- Il significato è acquisito e riconosciuto attraverso la pratica quotidiana del suo uso
  - *The meaning of a word is to be defined by the rules for its use, not by the feeling that attaches to the words*

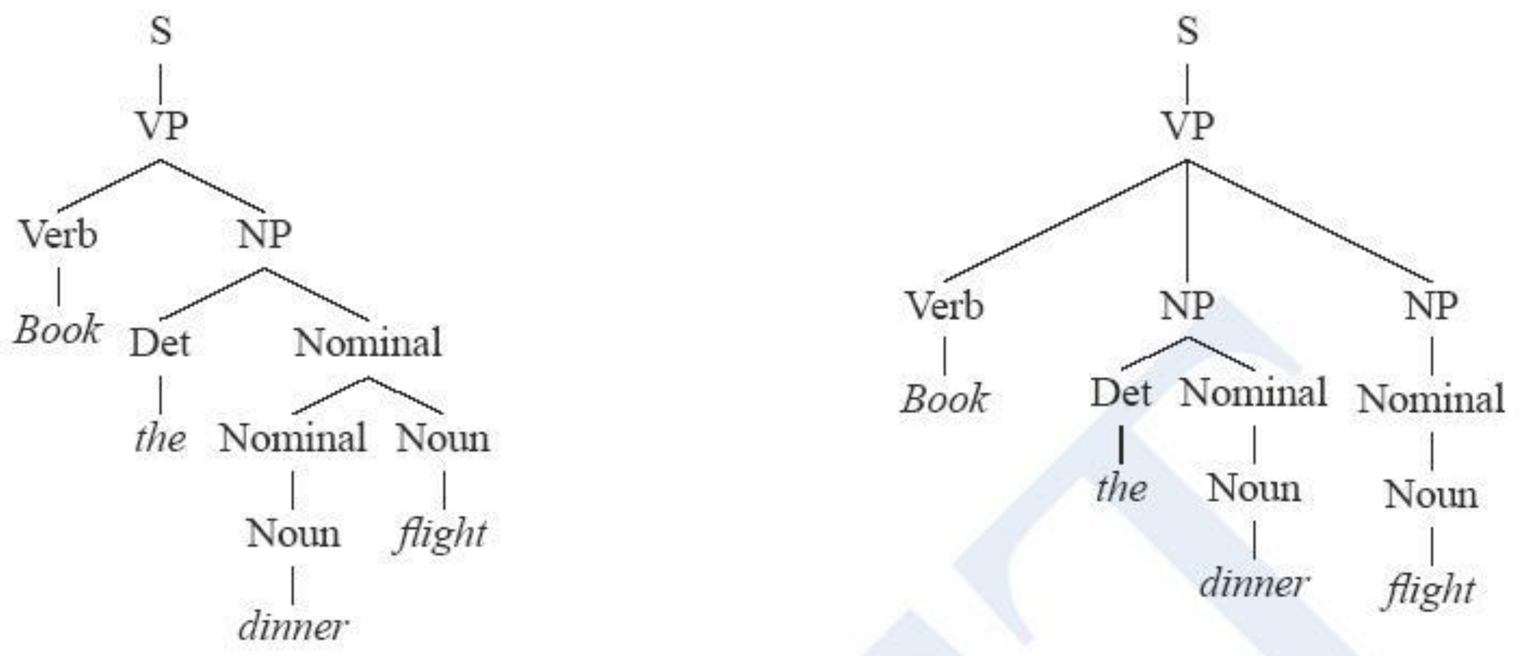
    L. Wittgenstein's Lectures, Cambridge 1932-1935.

- Comprendere un significato consiste nel collegare una espressione linguistica ad una esperienza (*praxis*) attraverso meccanismi quali la **analogia** o la limitazione del **rischio di essere fraintesi/inappropriati/oscuri**
- La **interpretazione** è riconducibile alla induzione di una o più **funzioni di decisione dalla esperienza**

Dal ragionamento simbolico alle rappresentazioni quantitative ed alla inferenza induttiva

Machine Learning: *weapons*

- Rule and Pattern learning from Data
  - *Frequent Pattern Mining* (*Basket analysis*)
- Probabilistic Extensions of Natural Language Grammars
  - Probabilstic CFGs
  - Stochastic Grammars
- Bayesian Models & Graphical Models
- Learning Discriminativo nei percettroni e reti neurali
- Speciali tipi di percettroni: le SVM
  - Funzioni Kernel in spazi di rappresentazione impliciti
- Il potere dell'ottimizzazione: il *deep learning*

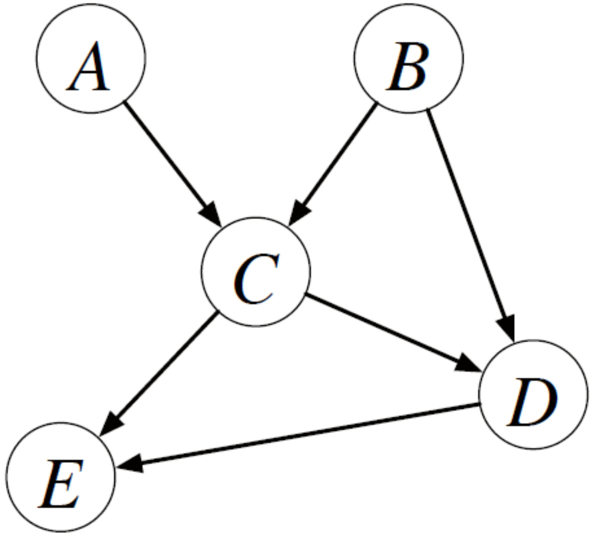


Rules			P
S	→	VP	.05
VP	→	Verb NP	.20
NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20
Nominal	→	Noun	.75
Verb	→	book	.30
Det	→	the	.60
Noun	→	dinner	.10
Noun	→	flights	.40

Rules			P
S	→	VP	.05
VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20
NP	→	Nominal	.15
Nominal	→	Noun	.75
Nominal	→	Noun	.75
Verb	→	book	.30
Det	→	the	.60
Noun	→	dinner	.10
Noun	→	flights	.40

Figure 13.2 Two parse trees for an ambiguous sentence. The transitive parse (a) cor-

Modelli Grafici Bayesiani



$$p(A, B, C, D, E) = p(A)p(B)p(C|A, B)p(D|B, C)p(E|C, D)$$

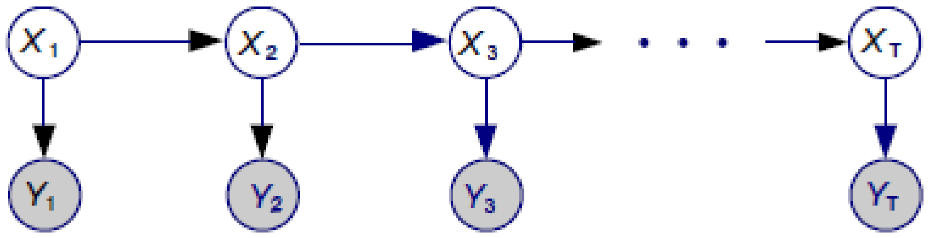
Lingue come Modelli Markoviani

- Descrivono un linguaggio in termini di un processo markoviano che, tramite un memoria limitata, esprime la probabilità di ogni sequenza
- Applicazioni:
  - Elaborazione del parlato
  - Riconoscimento della lingua
  - Machine Translation
  - Word Embeddings (Lexical Semantic)

$$p(X_{1,\ldots,T}) = p(X_1)\prod_{t=2}^{T} p(X_t|X_{t-1})$$

Hidden Markov Models (HMM)

- Stati (nascosti) = Categorie/Classi
- Osservazioni

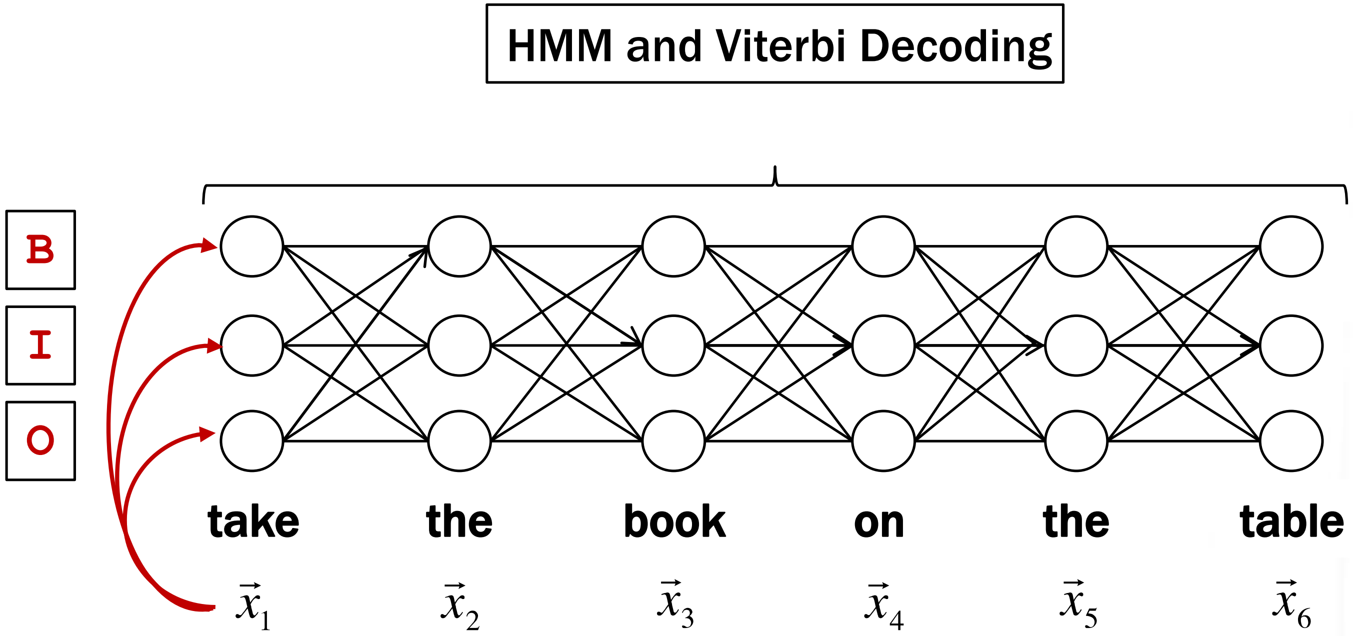


$$p(X_{1,\ldots,T}, Y_{1,\ldots,T}) = p(X_1)p(Y_1|X_1)\prod_{t=2}^{T}[p(X_t|X_{t-1})p(Y_t|X_t)]$$

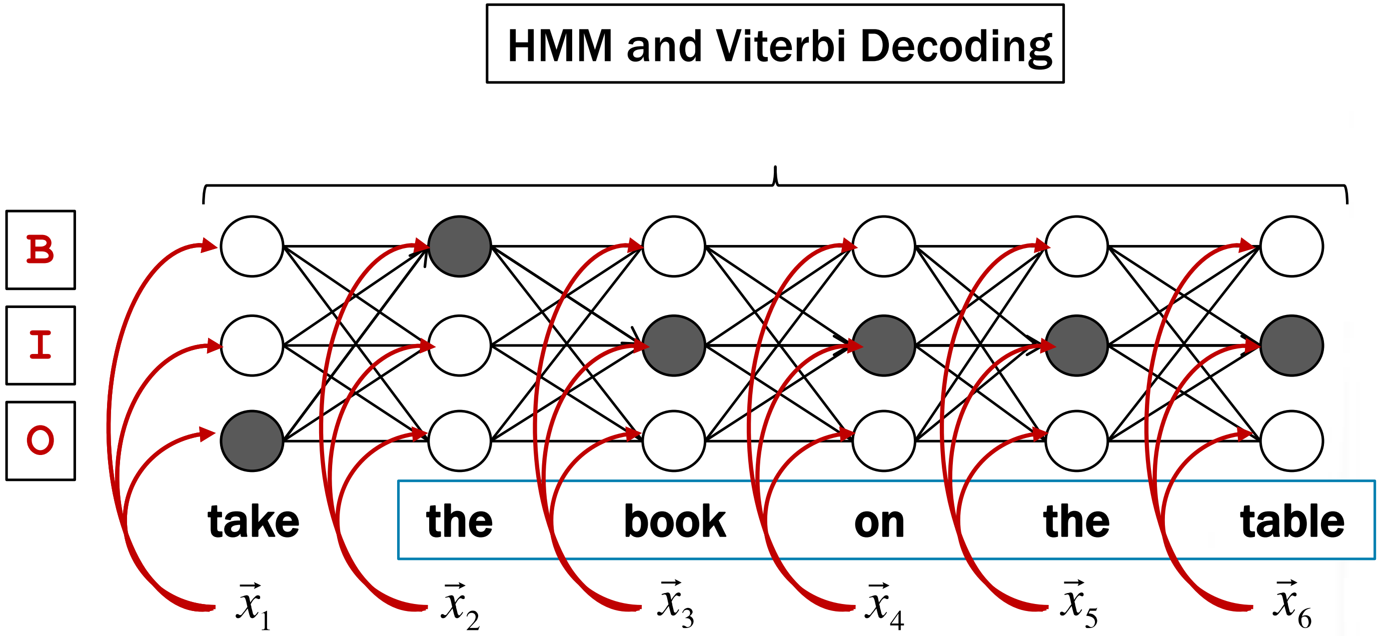
- Emissioni di simboli
- Transizioni (tra Stati)

- Applicazioni
  - Recognition del parlato
  - Task di *Sequence Labeling* (e.g. *POS tagging*)

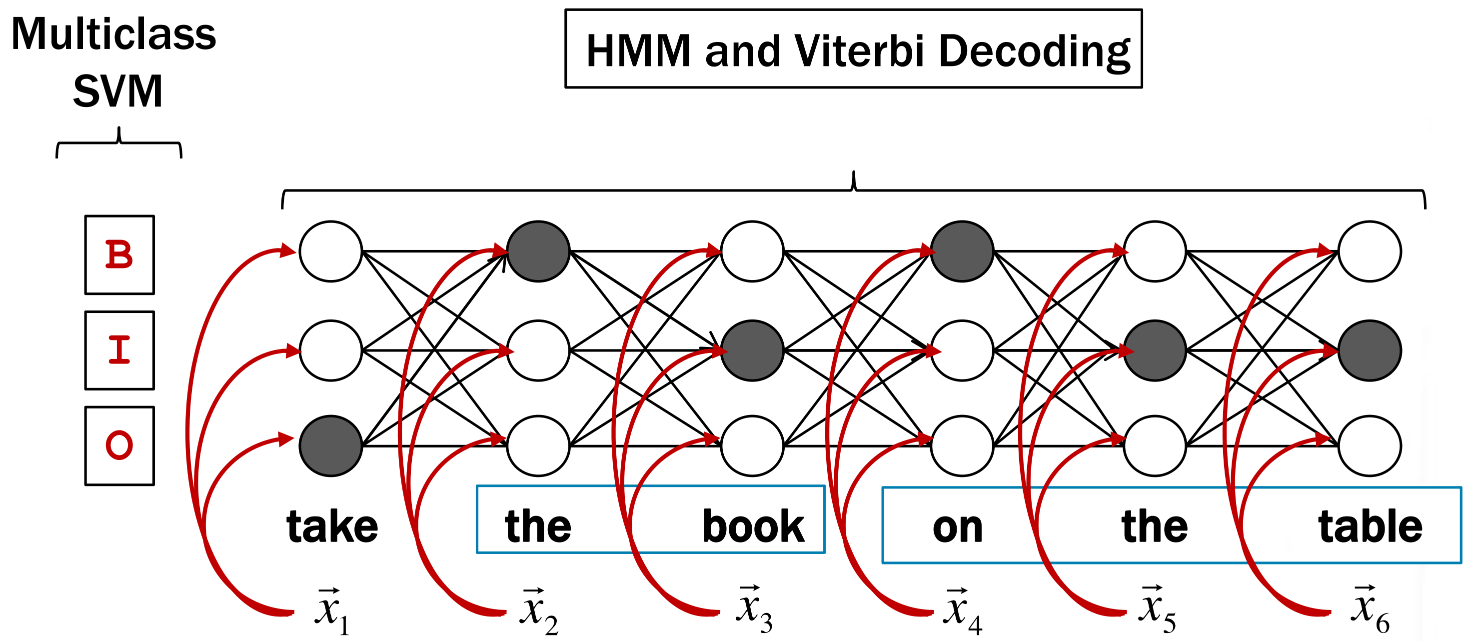
HMM & sequence labeling



HMM & sequence labeling



HMM & sequence labeling



HMM decoding in NLP

- Molti task linguistici sono riconducibili ad un processo di HMM sequence labeling :
  - Tokenizzazione
  - Riconoscimento di Multi Word Expression
  - POS tagging
  - Riconoscimento e Classificazione delle Entità
  - Semantic Parsing
  - Shallow Semantic Parsing over Framenet

Multiword Expressions

he was willing to budge a little on

O O O O B b i I

the price which means a lot to me .

O O O B I I I I O

a little; *means a lot to me*; *budge . . . on*

See: "Discriminative lexical semantic segmentation with gaps: running the MWE gamut," Schneider et al. (2014).

Named Entity Recognition

With Commander Chris Ferguson at the helm ,

person

O B I I O O O O

Atlantis touched down at Kennedy Space Center .

spacecraft location

B O O O B I I O

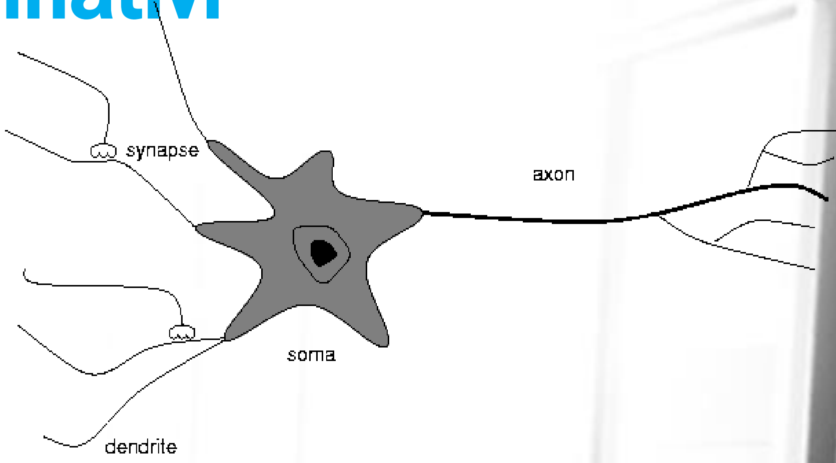
Supersense Tagging

ikr	smh	he	asked	fir	yo	last	name
–	–	–	communication	–	–	–	cognition

so	he	can	add	u	on	fb	lololol
–	–	–	stative	–	–	group	–

See: "Coarse lexical semantic annotation with supersenses: an Arabic case study," Schneider et al. (2012).

Machine Learning: Classificatori discriminativi



In uno spazio *n*-dimensionale, la funzione :

$$f(\vec{x}) = \vec{x} \cdot \vec{w} + b, \quad \vec{x}, \vec{w} \in \Re^n, b \in \Re$$

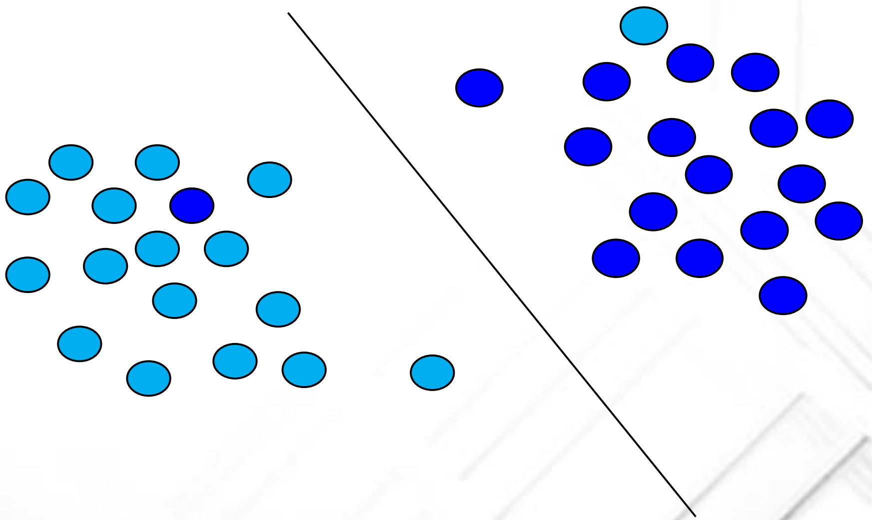
dove:

$\vec{x}$ è il vettore che descrive un esempio in input

$\vec{w}$ è il gradiente di un iperpiano

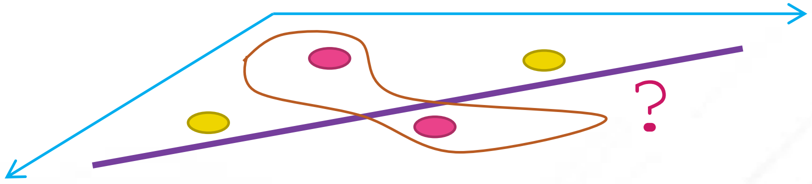
corrisponde alla funzione caratteristica di un concetto C da apprendere

Classificazione: $$h(x) = \text{sign}(f(x))$$



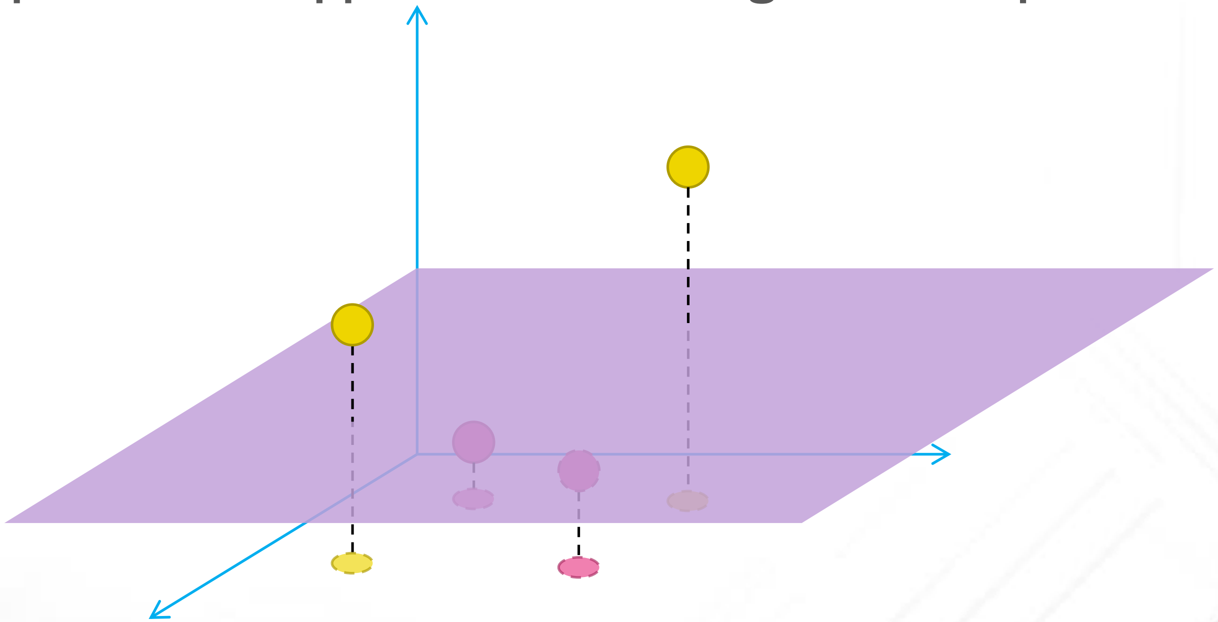
Apprendere rappresentazioni per concetti complessi

- Concetti complessi richiedono spazi complessi e funzioni complesse:
- Il concetto può non essere esprimibile correttamente da una legge lineare
  - In R^2, 4 punti non possono essere sempre separati (as VC=3)
    [Vapnik and Chervonenkis(1971)]
- Soluzione 1 (*neural networks*): Complessificare la funzione di classificazione
  - Architetture complesse basate su *ensemble di neuroni* (ad es. Reti multistrato)
  - Rischio di *over-fitting*



Kernel machines

- Soluzione 2: Mappare il problema in uno spazio a maggiore dimensionalità, i.e. un nuovo spazio delle proprietà attraverso una opportune funzione di proiezione φ
- Idea (from SLT): Spazi più complessi costituiscono un limite all'overfitting
- Gli spazi corrispondono a rappresentazioni adeguate di un problema

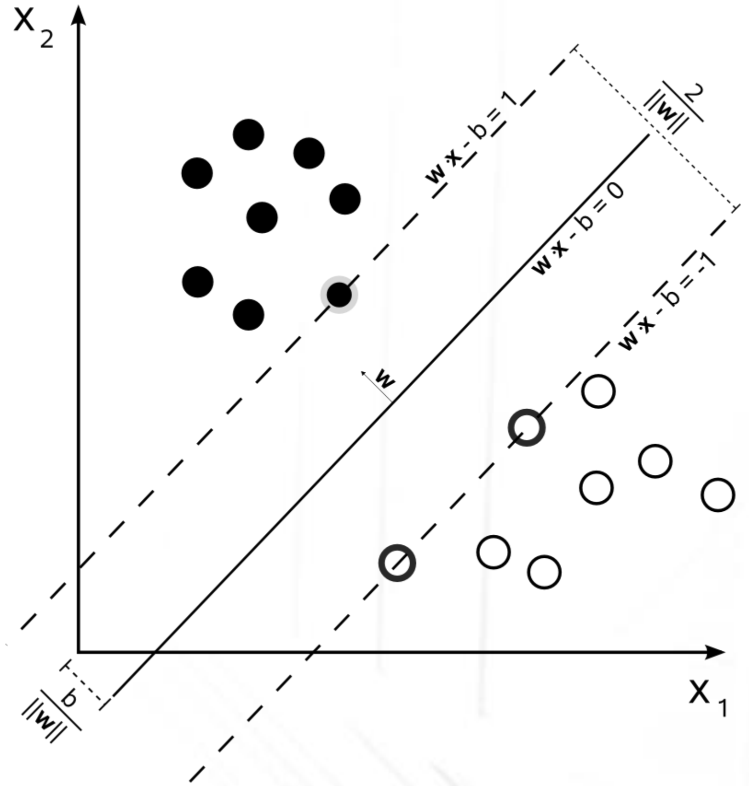


Representazione & Kernel

- Poichè la funzione lineare dipende dal prodotto scalare, i.e.

  $k(x_i,x_j)=\phi(x_i) \cdot \phi(x_j)$,

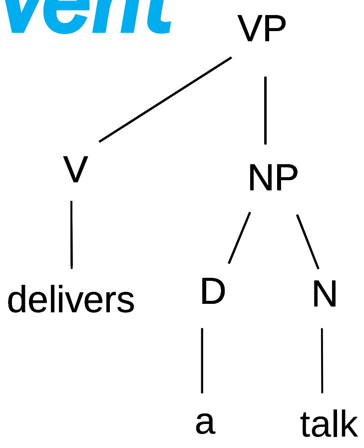
  per apprendere *h()* non è necessaria alcuna rappresentazione esplicita del vettore prodotta dalla funzione ϕ (Cristianini et al., 2002)
- Le SVM possono essere addestrate in spazi ad alta dimensionalità:
  - Disaccoppiando l'addestramento dalla scelta della rappresentazione ($\phi(x_j)$)
  - Esprimendo strutture complesse (ad es. alberi)
- *k(.,.)* corrisponde ad una metrica di similarità (e.g. legata al **lessico**, **sintassi** and/or **semantica**)



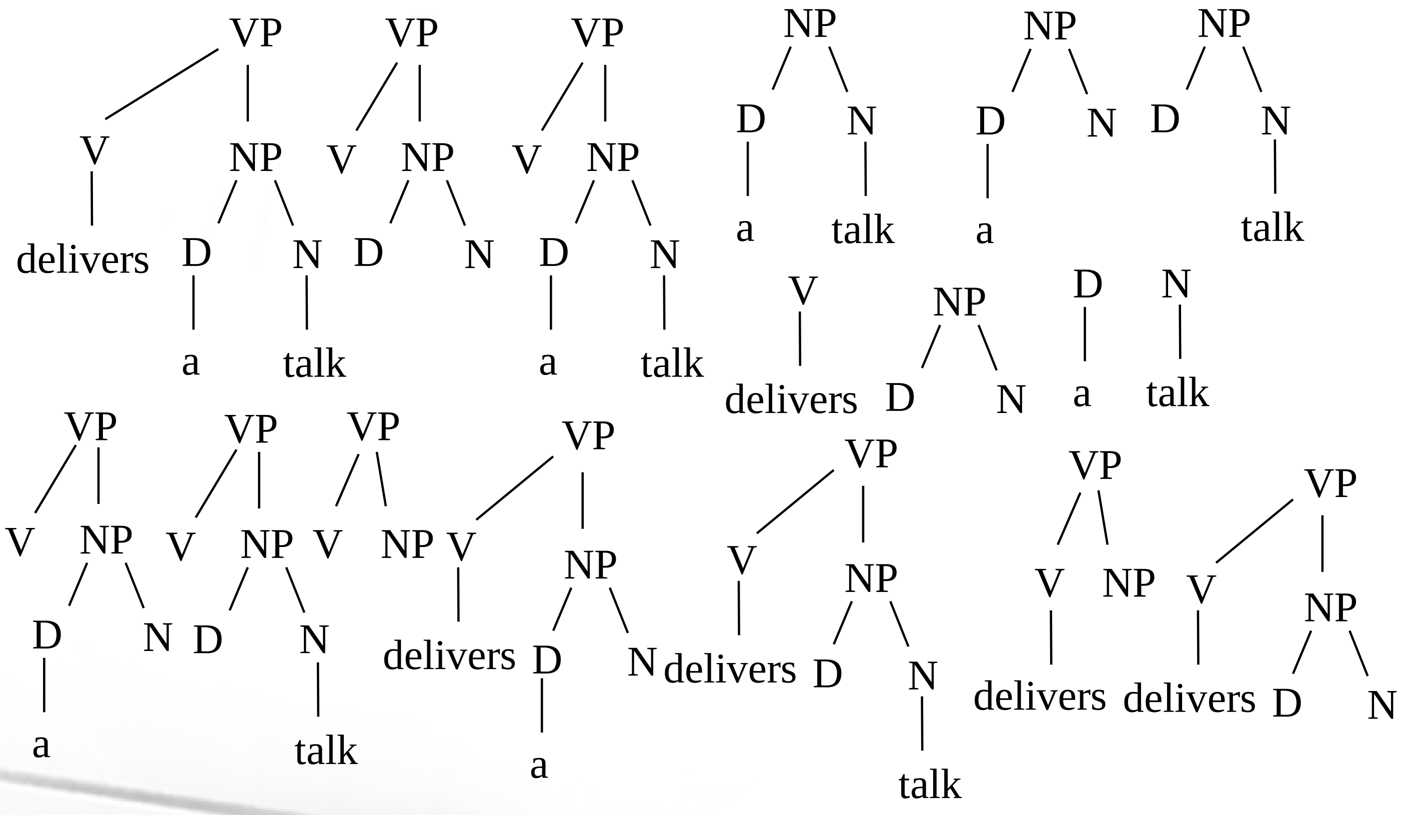
$$h(x) = \mathrm{sgn}(\vec{w}\cdot\varphi(\vec{x})+b) = \mathrm{sgn}(\underbrace{\sum_{j=1..\ell}\alpha_j y_j \varphi(\vec{x}_j)}_{\text{Support Vectors}}\cdot\varphi(\vec{x})+b) =$$

$$= \mathrm{sgn}(\sum_{i=1..\ell}\alpha_j y_j k(\vec{x}_j,\vec{x})+b)$$

Kernels & Strutture Sintattiche: *joint event*

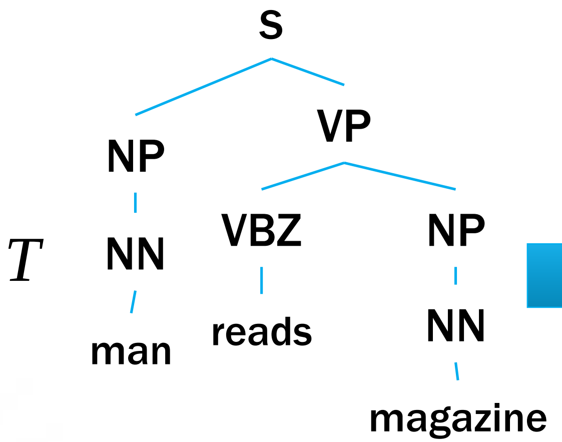


Un parse tree può essere visto come l'occorrere contemporaneo di un insieme di eventi (*joint event*) ....

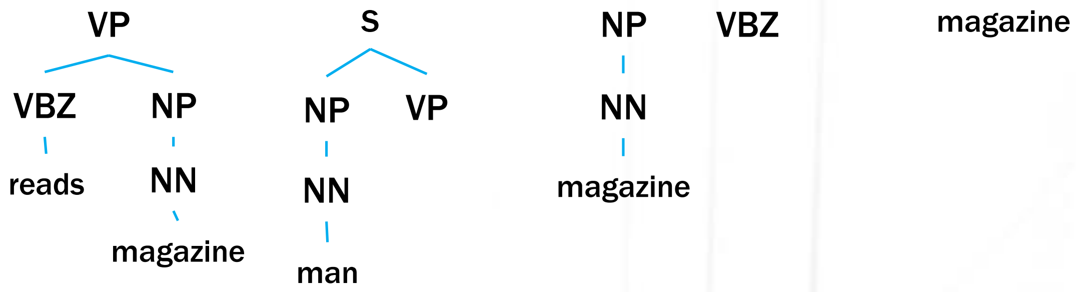


Dagli alberi allo spazio (implicito) Kernel

- Una funzione (implicita) φ trasforma un albero *T* in un vettore che rappresenta TUTTI i suoi sottoalberi, esprimendo
  - Informazione Lessicale (*man, magazine*)
  - Informazione grammaticale: POS tags e frammenti complessi dell'albero



$$\varphi(T) = (0,\ldots,1,\ldots,0,\ldots,1,\ldots,0,\ldots,1,\ldots,1,\ldots,0,\ldots,1)$$



- Il prodotto scalare $\varphi(T)\cdot\varphi(T')$ opera nello spazio di tutti i sottoalberi (possibili) ed è proporzionale al numero dei sottoalberi condivisi tra T e T'
- L'algoritmo neurale (e.g. SVM) seleziona gli esempi (le strutture salienti) che siano discriminative rispetto al task

Deep Learning

- "A physical symbol system has the necessary and sufficient means for general intelligent action.

  --Allen Newell &

  Herbert Simon

- Symbols are Luminiferous Aether of AI

  —Geoff Hinton

Multilayer Perceptrons

$$NN_{MLP2}(\mathbf{x}) = (g^2(g^1(\mathbf{x}\mathbf{W^1} + \mathbf{b^1})\mathbf{W^2} + \mathbf{b^2}))\mathbf{W^3}$$

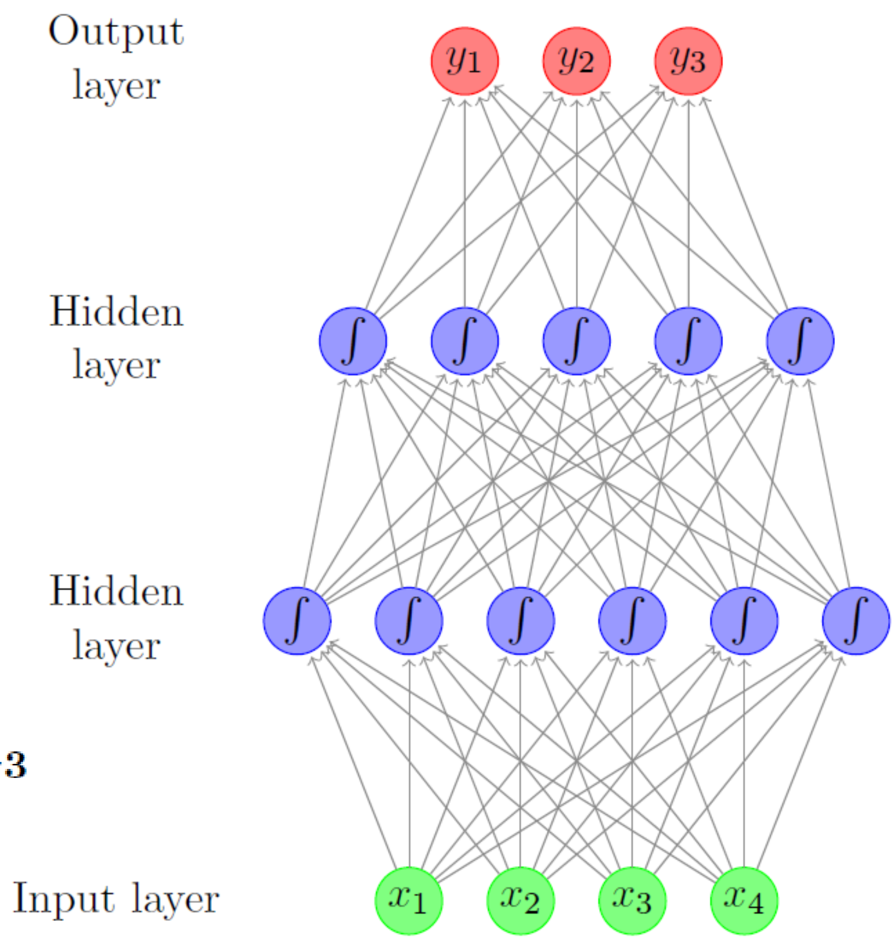
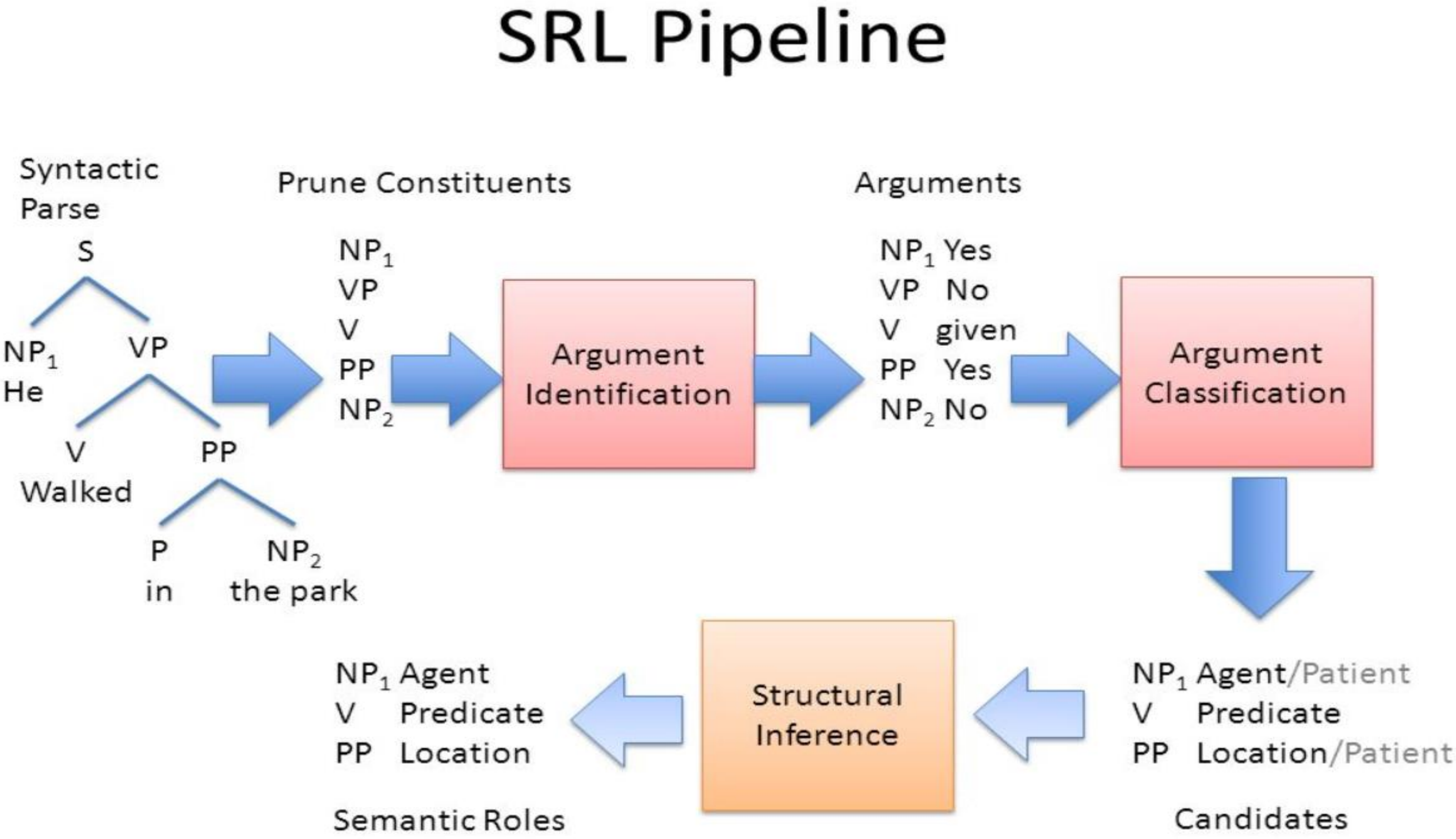


Figure 2: Feed-forward neural network with two hidden layers.

Semantic parsing su Framenet/PropBank



Recurrent Neural Networks

For example, consider the classical form of a dynamical system:

$$\boldsymbol{s}^{(t)} = f(\boldsymbol{s}^{(t-1)}; \boldsymbol{\theta}), \tag{10.1}$$

where $\boldsymbol{s}^{(t)}$ is called the state of the system.

Equation 10.1 is recurrent because the definition of $\boldsymbol{s}$ at time t refers back to the same definition at time $t-1$.

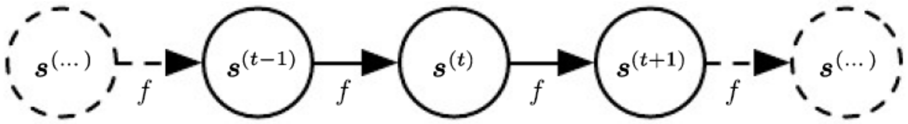
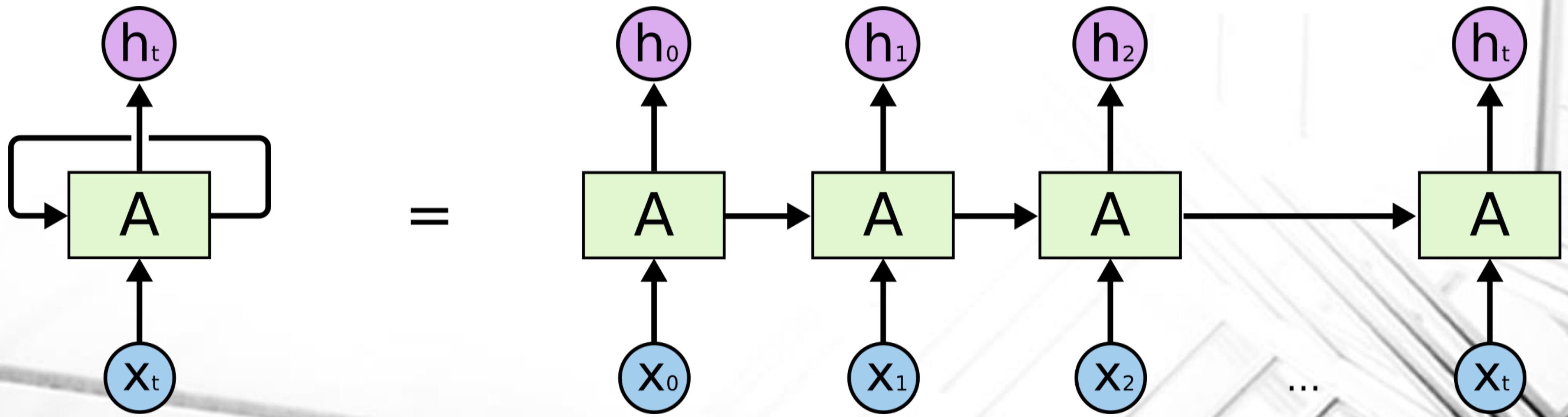


Figure 10.1: The classical dynamical system described by equation 10.1, illustrated as an unfolded computational graph. Each node represents the state at some time t, and the function f maps the state at t to the state at $t+1$. The same parameters (the same value of $\boldsymbol{\theta}$ used to parametrize f) are used for all time steps.



Training a RNN

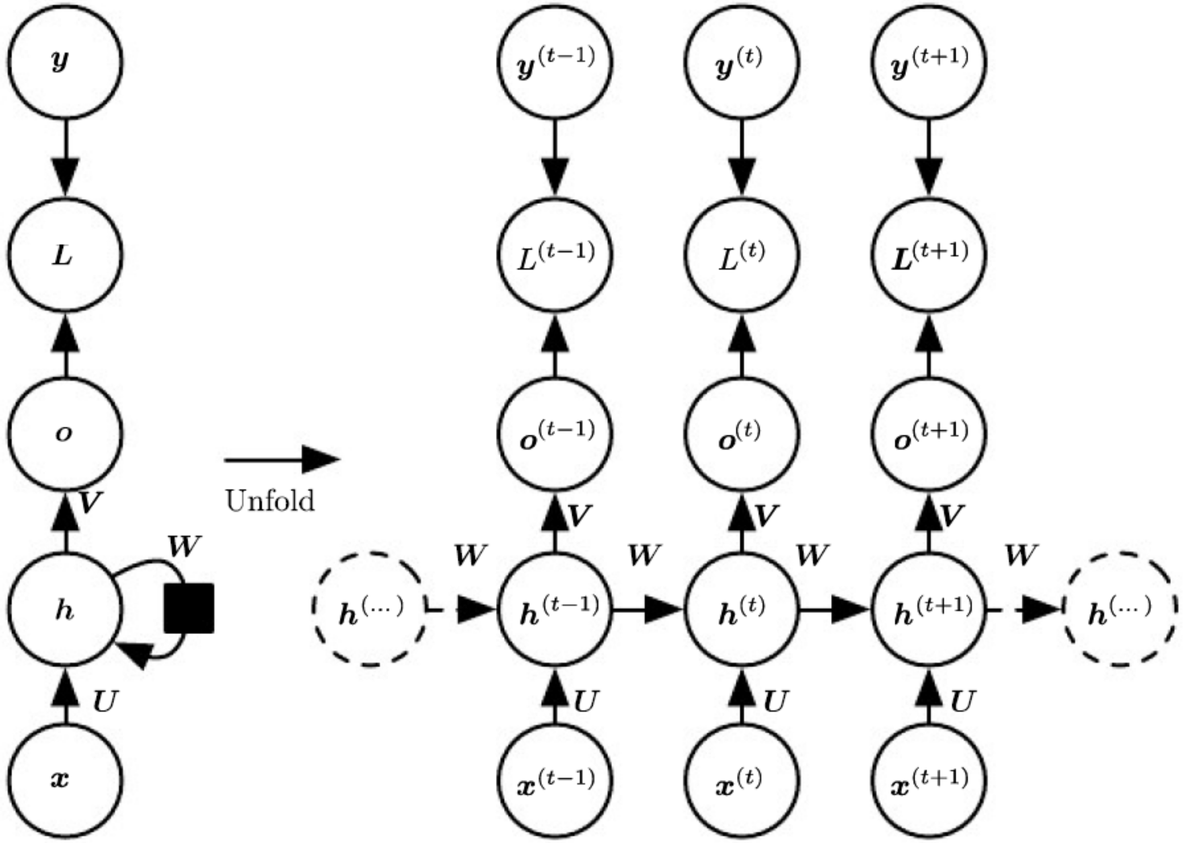


Figure 10.3: The computational graph to compute the training loss of a recurrent network that maps an input sequence of $\boldsymbol{x}$ values to a corresponding sequence of output $\boldsymbol{o}$ values. A loss L measures how far each $\boldsymbol{o}$ is from the corresponding training target $\boldsymbol{y}$. When using softmax outputs, we assume $\boldsymbol{o}$ is the unnormalized log probabilities. The loss L internally computes $\hat{\boldsymbol{y}} = \text{softmax}(\boldsymbol{o})$ and compares this to the target $\boldsymbol{y}$. The RNN has input to hidden connections parametrized by a weight matrix $\boldsymbol{U}$, hidden-to-hidden recurrent connections parametrized by a weight matrix $\boldsymbol{W}$, and hidden-to-output connections parametrized by a weight matrix $\boldsymbol{V}$. Equation 10.8 defines forward propagation in this model. *(Left)* The RNN and its loss drawn with recurrent connections. *(Right)* The same seen as a time-unfolded computational graph, where each node is now associated with one particular time instance.

Types of RNNs

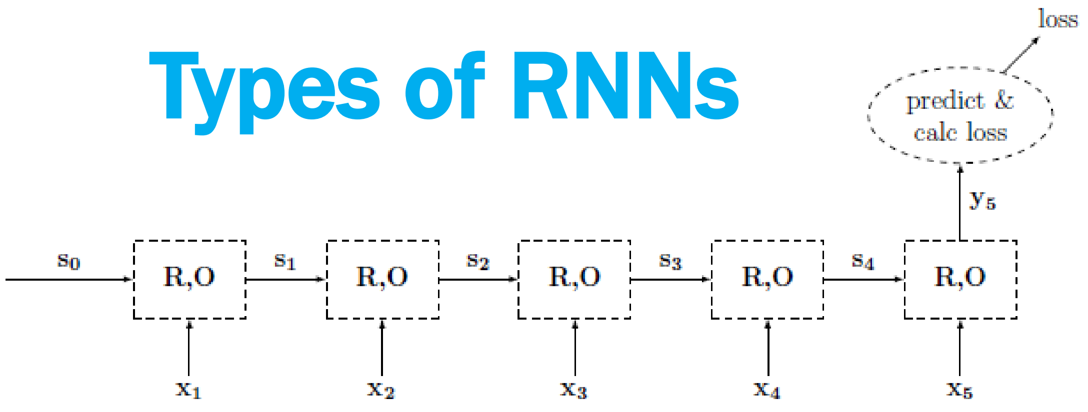


Figure 7: Acceptor RNN Training Graph.

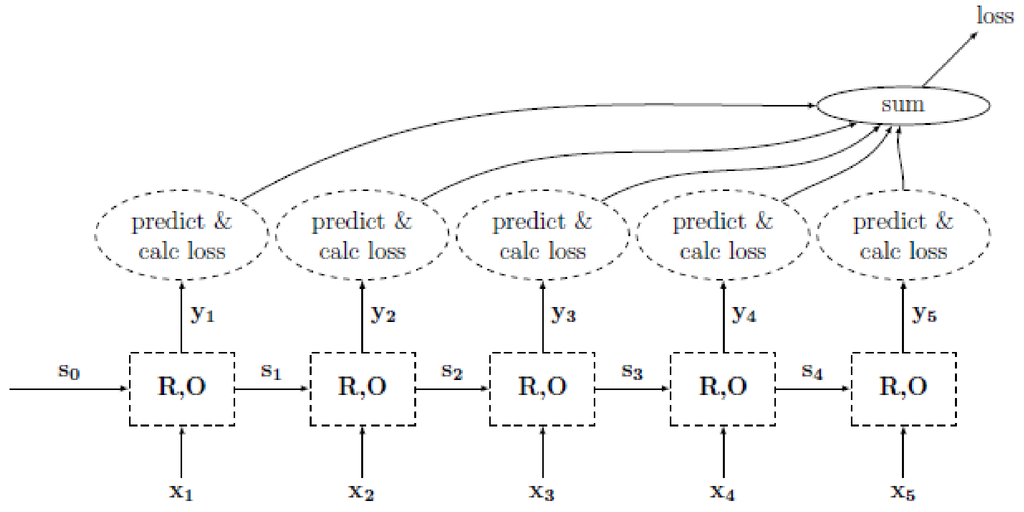


Figure 8: Transducer RNN Training Graph.

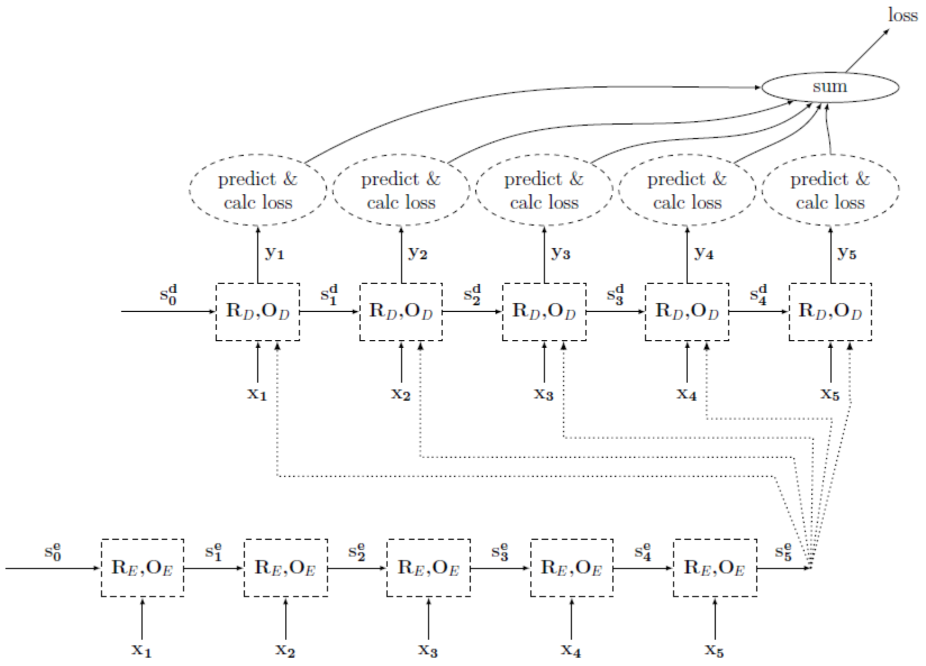


Figure 9: Encoder-Decoder RNN Training Graph.

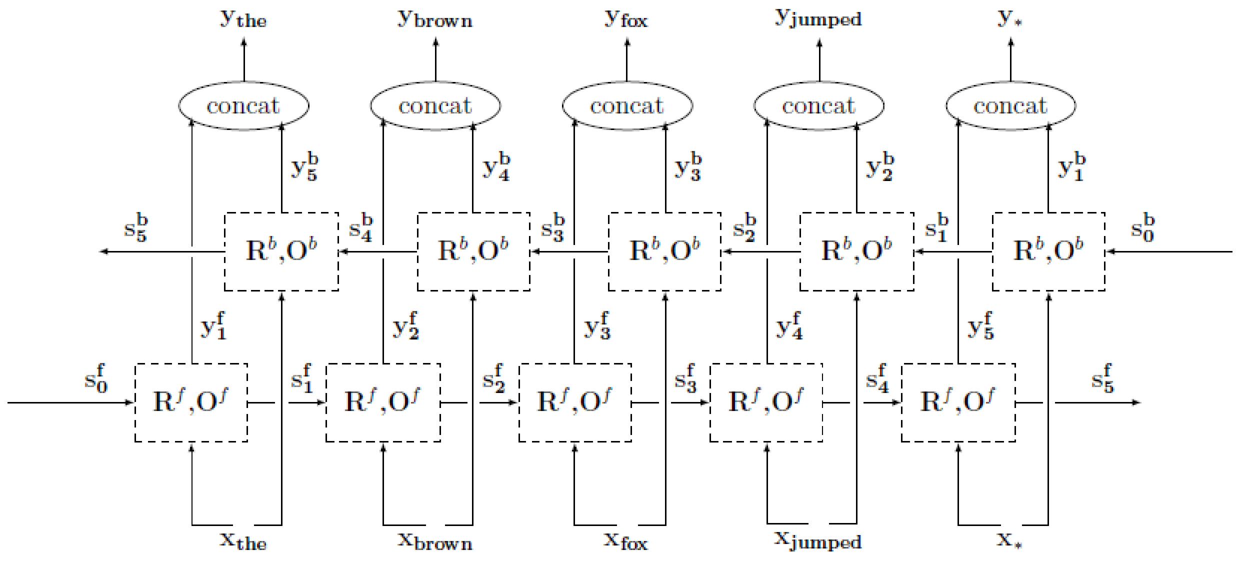
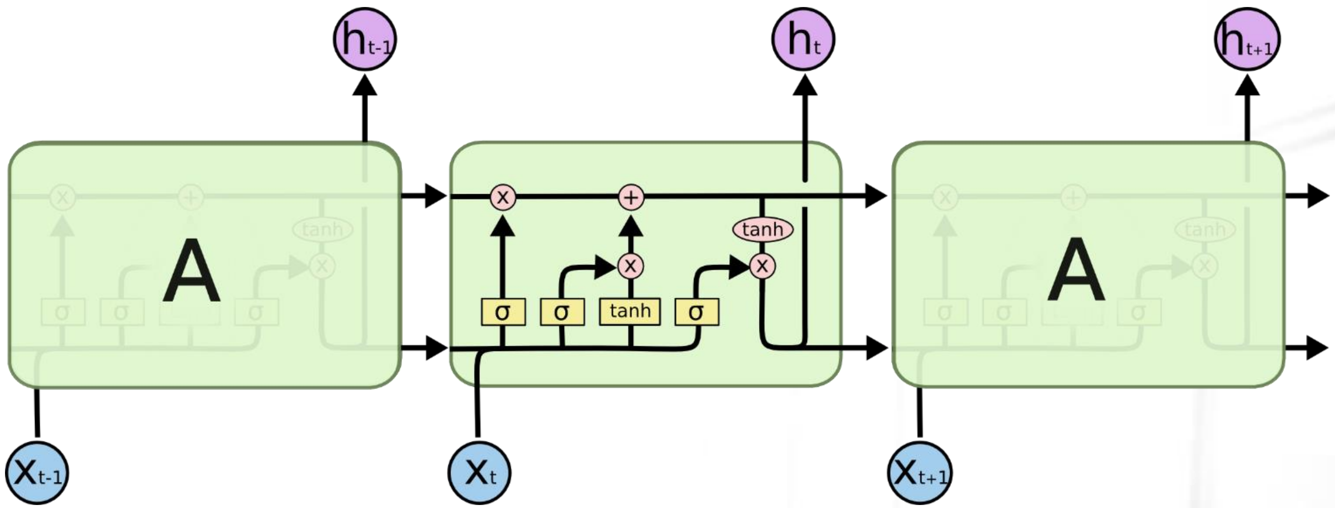


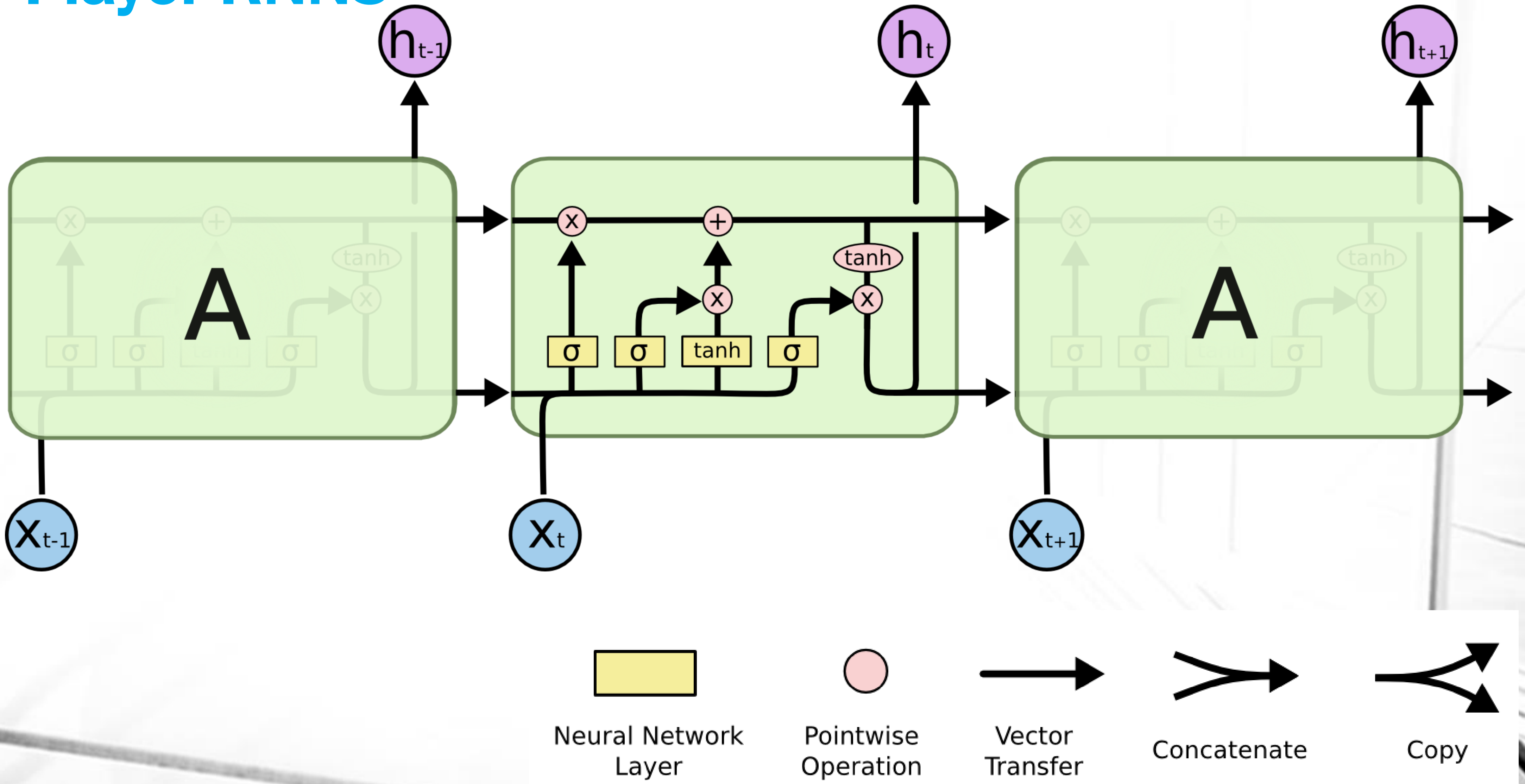
Figure 11: biRNN over the sentence “the brown fox jumped .”.

LSTMS



- At each input state, a gate is used to decide:
  - **how much of the new input** should be written to the memory cell,
  - **how much of the current content** of the memory cell should be **forgotten**.
- Concretely, a gate g in [0;1]n is a vector of values in the range [0; 1] that is multiplied component-wise with another vector *C* in R^n, and the result is then added to another vector.
- Indices in *C* corresponding to near-one values in *g* are allowed to pass, while those corresponding to near-zero values are blocked.

4 layer RNNS



h_{t-1}
h_t
h_{t+1}
A
A
σ
σ
tanh
σ
tanh
X_{t-1}
X_t
X_{t+1}
Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

Examples: Language understanding

https://github.com/Microsoft/CNTK/wiki/Hands-On-Labs-Language-Understanding

Task: Slot tagging with an LSTM

show	O
flights	O
from	O
burbank	B-fromloc.city_name
to	O
st.	B-toloc.city_name
louis	I-toloc.city_name
on	O
monday	B-depart_date.day_name

Microsoft

Examples: language understanding

https://github.com/Microsoft/CNTK/wiki/Hands-On-Labs-Language-Understanding

Task: Slot tagging with an LSTM

```
19   |x 178:1 |# BOS      |y 128:1 |# O
19   |x 770:1 |# show     |y 128:1 |# O
19   |x 429:1 |# flights  |y 128:1 |# O
19   |x 444:1 |# from     |y 128:1 |# O
19   |x 272:1 |# burbank  |y 48:1  |# B-fromloc.city_name
19   |x 851:1 |# to       |y 128:1 |# O
19   |x 789:1 |# st.      |y 78:1  |# B-toloc.city_name
19   |x 564:1 |# louis    |y 125:1 |# I-toloc.city_name
19   |x 654:1 |# on       |y 128:1 |# O
19   |x 601:1 |# monday   |y 26:1  |# B-depart_date.day_name
19   |x 179:1 |# EOS      |y 128:1 |# O
```

```
    ^
    |
+-------+
| Dense |
+-------+
    ^
    |
+------+
| LSTM |
+------+
    ^
    |
+-------+
| Embed |
+-------+
    ^
    |
```

Examples: language understanding

https://github.com/Microsoft/CNTK/wiki/Hands-On-Labs-Language-Understanding

Task: Slot tagging with an LSTM

```
19   |x 178:1 |# BOS       |y 128:1 |# O
19   |x 770:1 |# show      |y 128:1 |# O
19   |x 429:1 |# flights   |y 128:1 |# O
19   |x 444:1 |# from      |y 128:1 |# O
19   |x 272:1 |# burbank   |y 48:1  |# B-fromloc.city_name
19   |x 851:1 |# to        |y 128:1 |# O
19   |x 789:1 |# st.       |y 78:1  |# B-toloc.city_name
19   |x 564:1 |# louis     |y 125:1 |# I-toloc.city_name
19   |x 654:1 |# on        |y 128:1 |# O
19   |x 601:1 |# monday    |y 26:1  |# B-depart_date.day_name
19   |x 179:1 |# EOS       |y 128:1 |# O
```

```
y         "O"         "O"         "O"         "O"   "B-fromloc.city_name"
           ^           ^           ^           ^           ^
           |           |           |           |           |
       +-------+   +-------+   +-------+   +-------+   +-------+
       | Dense |   | Dense |   | Dense |   | Dense |   | Dense |  ...
       +-------+   +-------+   +-------+   +-------+   +-------+
           ^           ^           ^           ^           ^
           |           |           |           |           |
       +------+    +------+    +------+    +------+    +------+
0 -->| LSTM |-->| LSTM |-->| LSTM |-->| LSTM |-->| LSTM |-->...
       +------+    +------+    +------+    +------+    +------+
           ^           ^           ^           ^           ^
           |           |           |           |           |
       +-------+   +-------+   +-------+   +-------+   +-------+
       | Embed |   | Embed |   | Embed |   | Embed |   | Embed |  ...
       +-------+   +-------+   +-------+   +-------+   +-------+
           ^           ^           ^           ^           ^
           |           |           |           |           |
x ------>+--------->+--------->+--------->+--------->+------...
          BOS      "show"    "flights"    "from"    "burbank"
```

Bi-directional RRNs (Schuster and Paliwal, 1997)

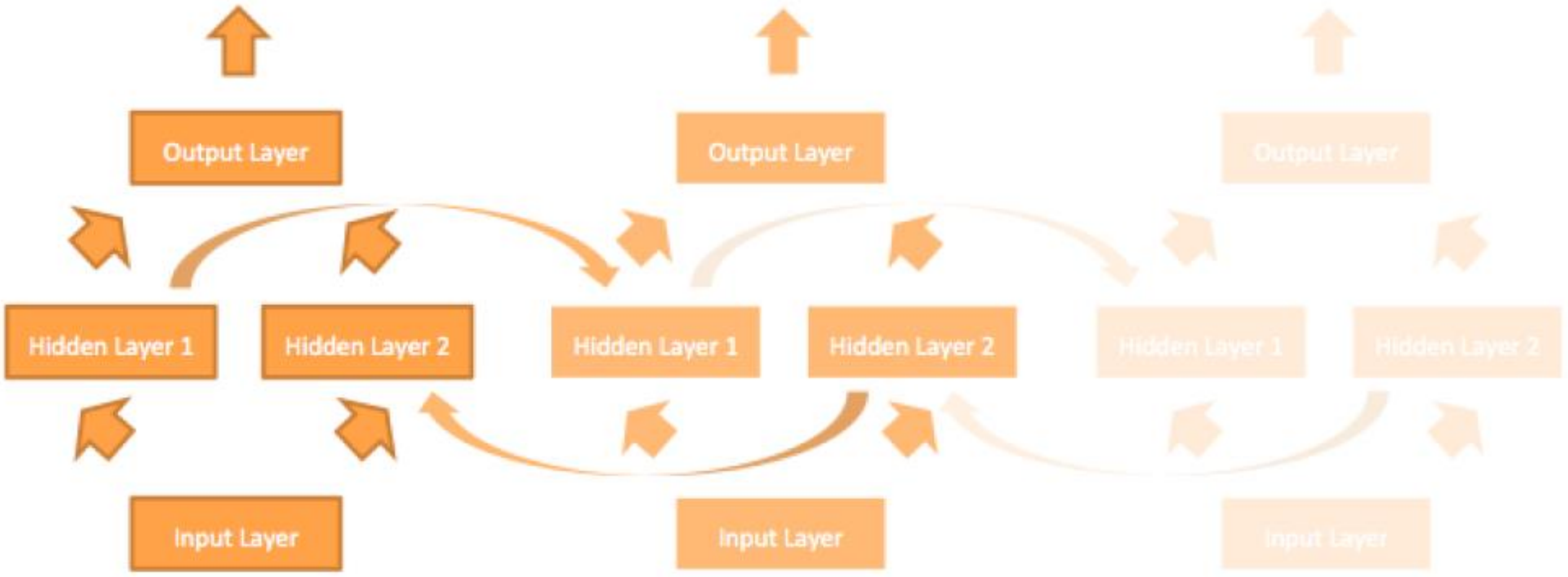
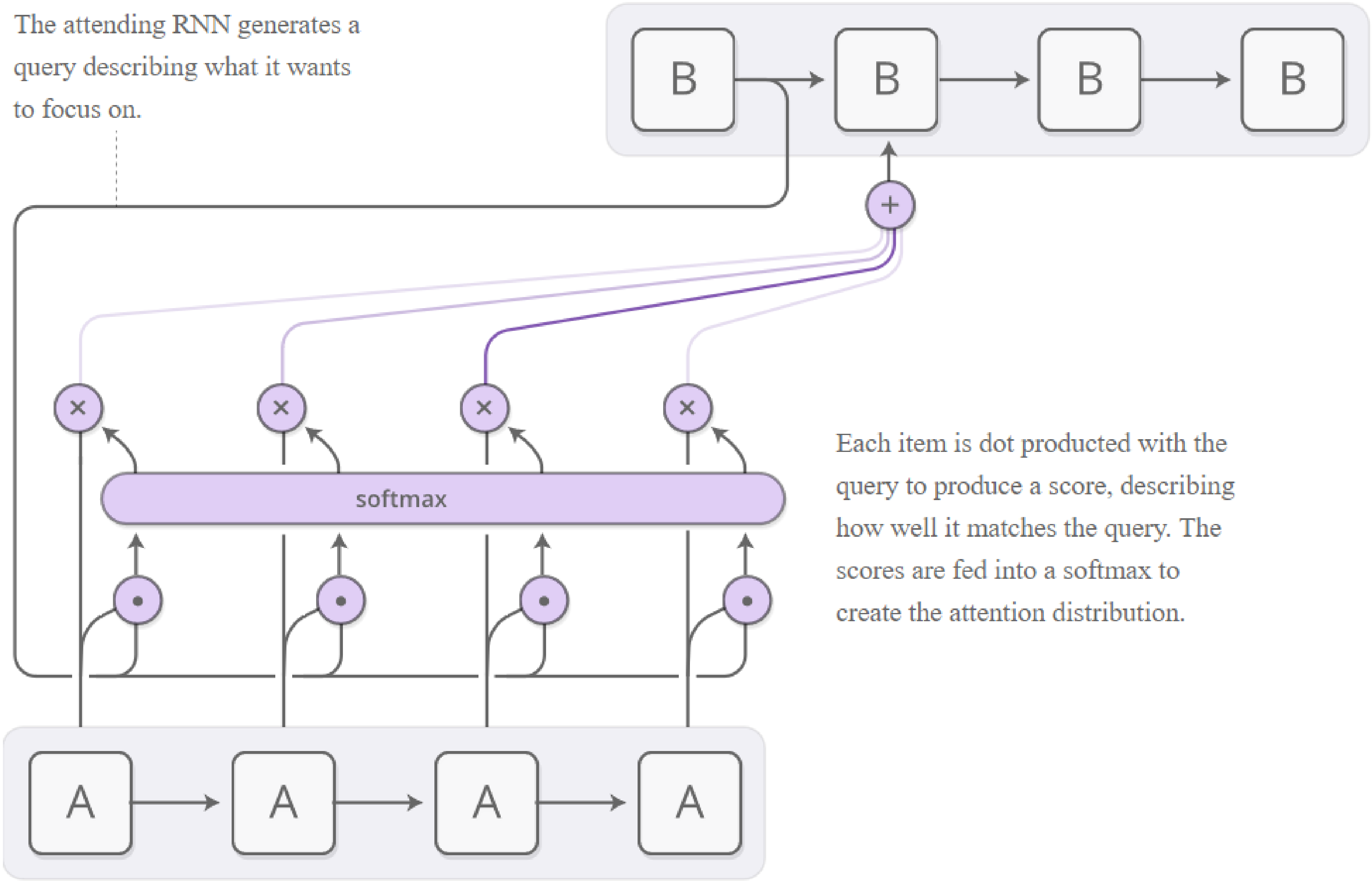


Figure 23: The unfolded structured of BRNN. The temporal order is from left to right. Hidden layer 1 is unfolded in the standard way of an RNN. Hidden layer 2 is unfolded to simulate the reverse connection.

Attention-based RNNs



Attention mechanisms in Machine Translation

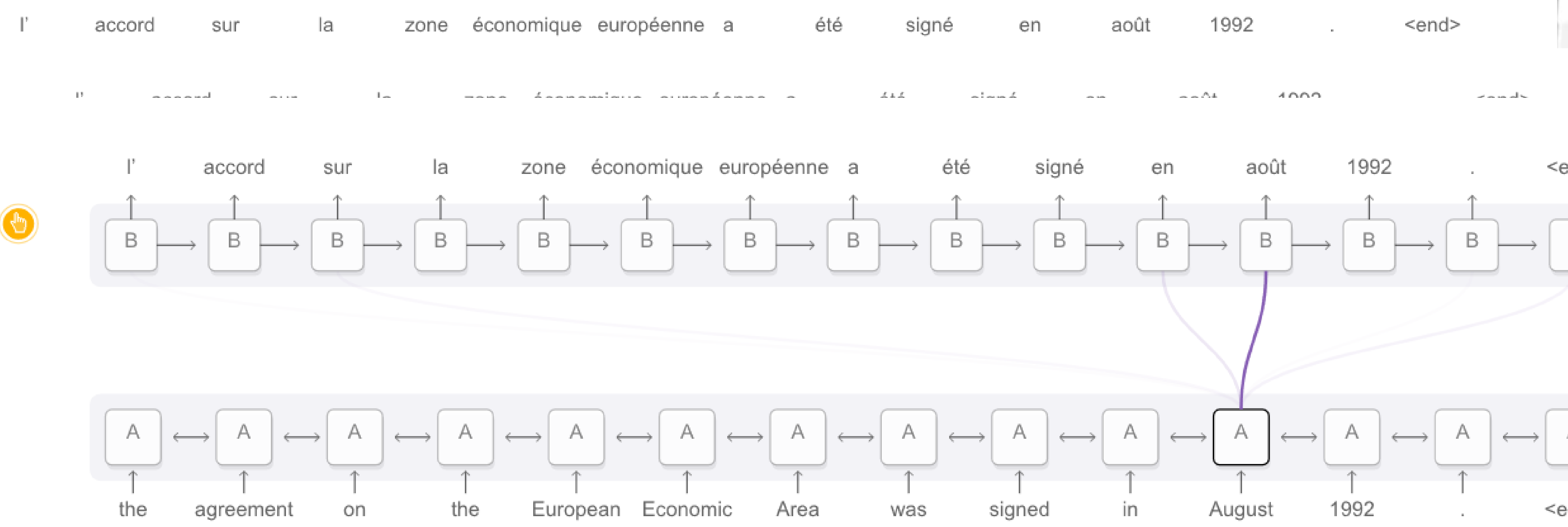


Diagram derived from Fig. 3 of Bahdanau, *et al.* 2014

WSD with Bi-LSTM (Raganato et al., EMNLP 2017)

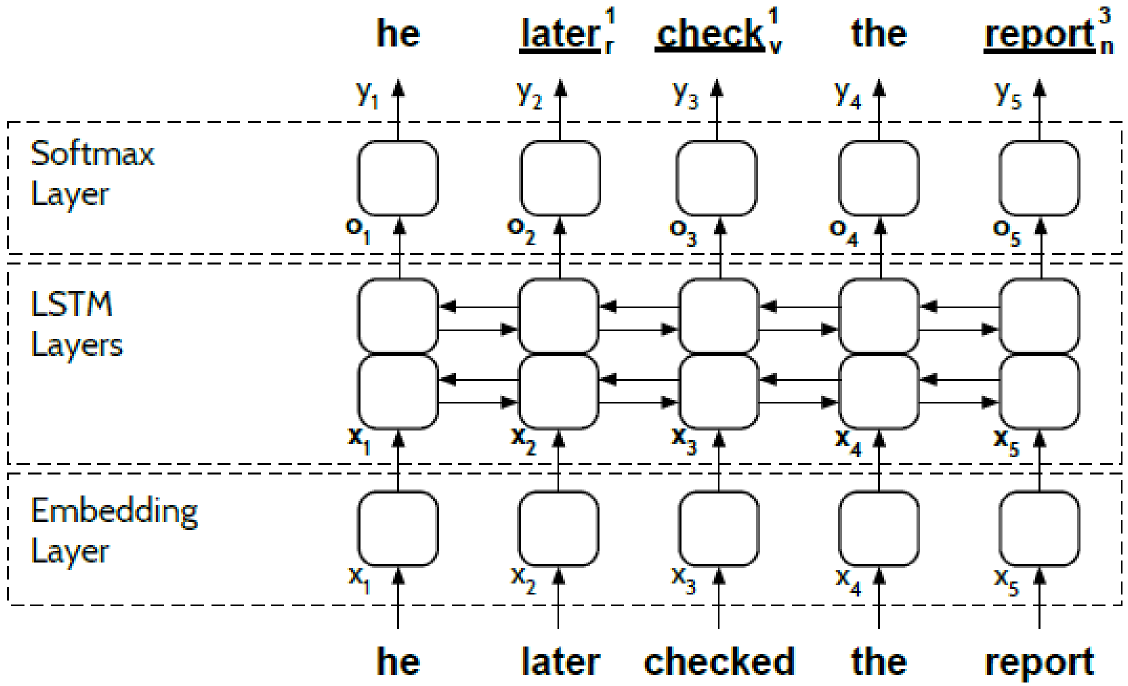


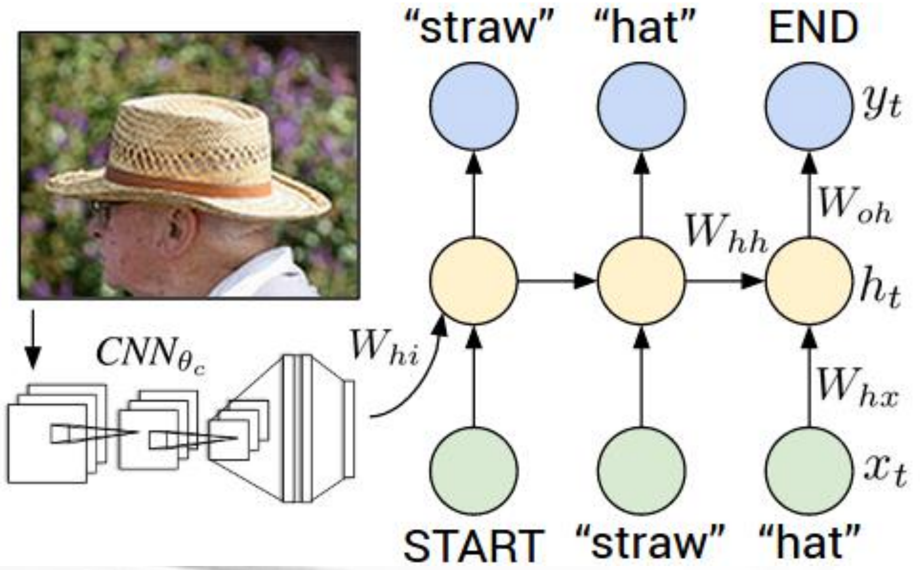
Figure 1: Bidirectional LSTM sequence labeling architecture for WSD (2 hidden layers). We use the notation of Navigli (2009) for word senses: w_p^i is the i-th sense of w with part of speech p.

A complex application: Image captioning

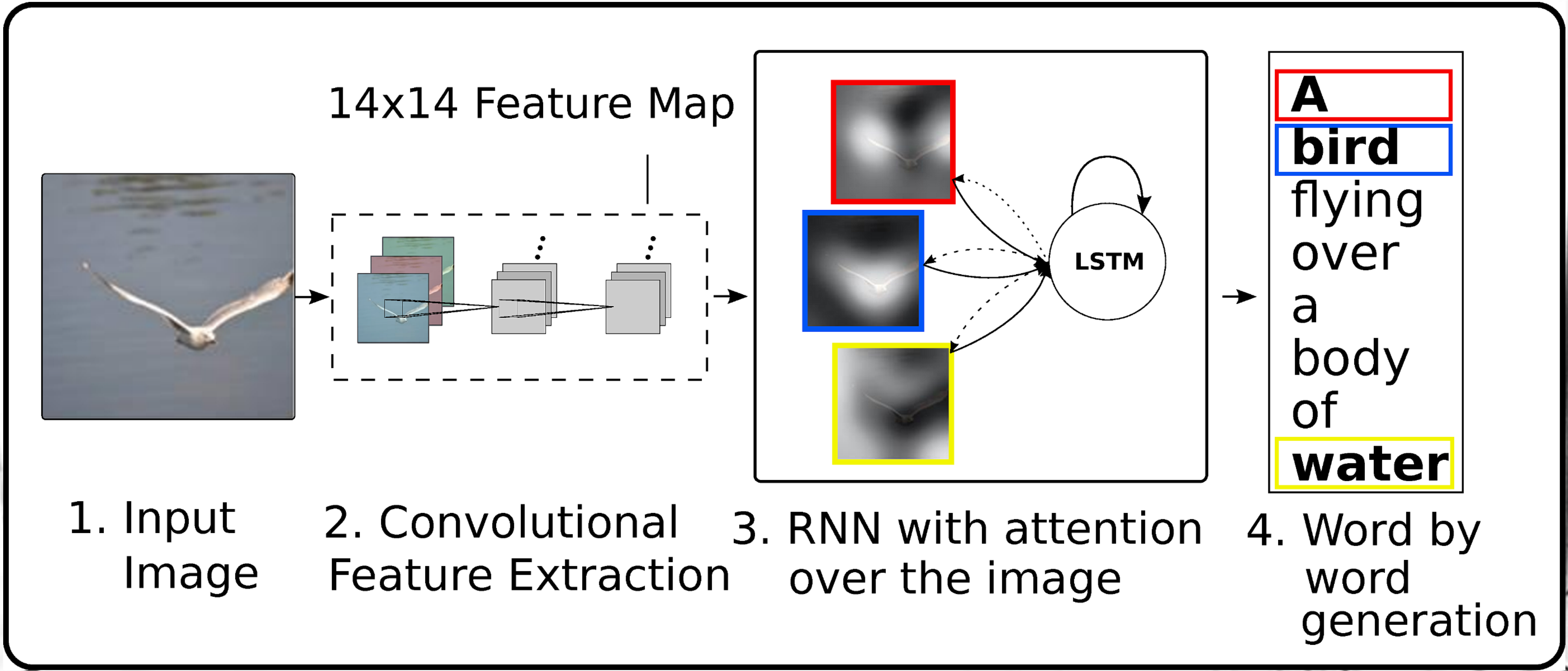
A woman is throwing a frisbee in a park. A dog is standing on a hardwood floor. A stop sign is on a road with a mountain in the background.

Other Advanced architectures

- Image to captions
  - Convolutional Neural Network to learn a representation of the image
  - (Bi-directional) Recurrent Neural Network to generate a caption describing the image
    - its input is the representation computed from the CNN
    - its output is a sequence of words, i.e. the caption



"baseball player is throwing ball in game."


14x14 Feature Map
LSTM
A
bird
flying
over
a
body
of
water
1. Input Image
2. Convolutional Feature Extraction
3. RNN with attention over the image
4. Word by word generation

A(0.99)

A(1.00)

Show & Tell in italiano

- Current work at UniTV (Croce, Masotti, Basili, 2017)

(a) im2txt+translation: *Un giocatore di baseball che oscilla una mazza ad una sfera*, Italian model: *Un giocatore di baseball che tiene una mazza da baseball su un campo.*

(b) im2txt+translation: *Una grande torre dell'orologio che sovrasta una città*, Italian model: *Un grande edificio con un orologio sulla parte superiore.*

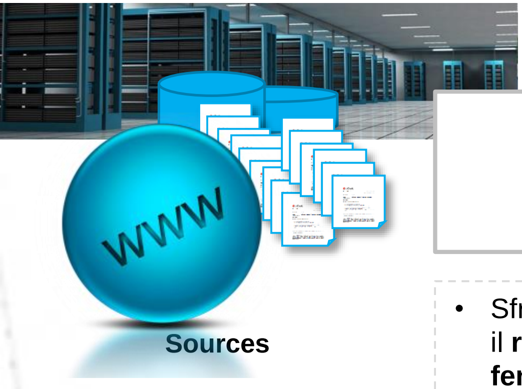
(c) im
di per
dei c
che c

(d) im2txt+translation: *Una persona che salta una tavola skate in aria*, Italian model: *Un uomo che cavalca uno skateboard su una strada.*

Outline

- **Artificial Intelligence & Natural Language Processing**
  - Comunicazione linguistica & Conoscenza
  - Il ruolo dei dati
- **Natural Language Processing: *Task*, Modelli e Metodi**
- **Trattamento delle lingue e *Machine Learning***
  - Statistical Language Processing
  - Apprendimento discriminativo per l'NLP
- **Natural Language Processing: applications**
- **Conclusions & Perspectives**

Applicazioni NLP: a roadmap



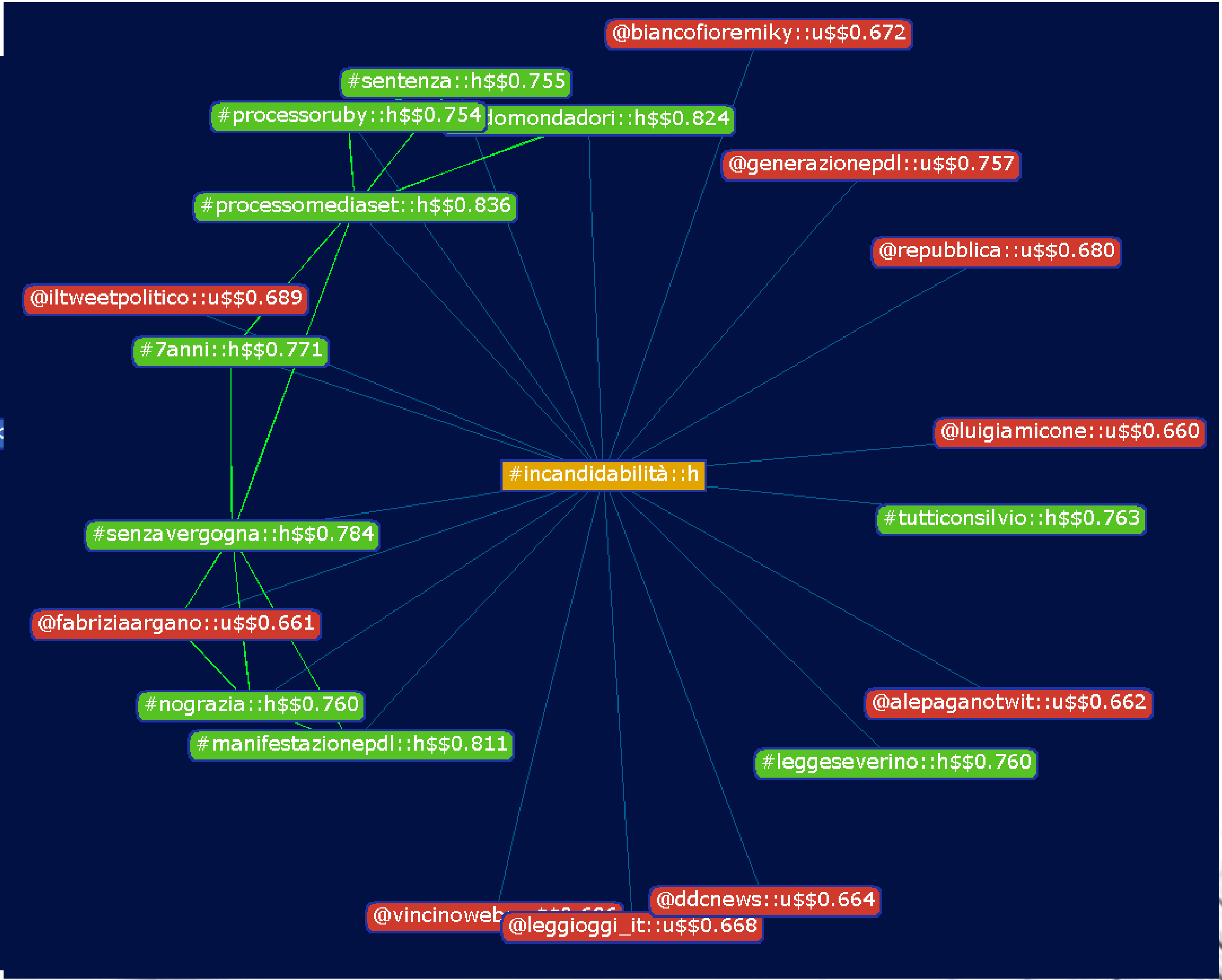
Conoscenza Operativa,
Verifica dei Fatti & Trust *checking*

NLP sui testi	Concettualizzazione	Esplorazione e Predizione
• Sfrutta modelli linguistici per il **riconoscimento di fenomeni semantici** • **Risolvi** le principali **ambiguità di senso** • **Scala al trattamento di tutti i documenti coinvolti** (grandi archivi documentali e il Web)	• **Riconoscimento di fenomeni impliciti** • **Analisi in ampiezza delle fonti documentali** e ***fact checking*** a grana fine • Scoperta di **nuovi fatti di interesse (globale)**	• **Verifica della consistenza dei fatti individuali** • **Aggregazione di fatti correlati** • **Validazione empirica di nuove ipotesi interpretative** • **Planning di nuove analisi**

Use cases

- Distributional Semantics (Embeddings vettoriali)
- Semantic Search engines
- NLP in supporto alla Investigazione
- Sentiment analysis over twitter
- Psychological modeling

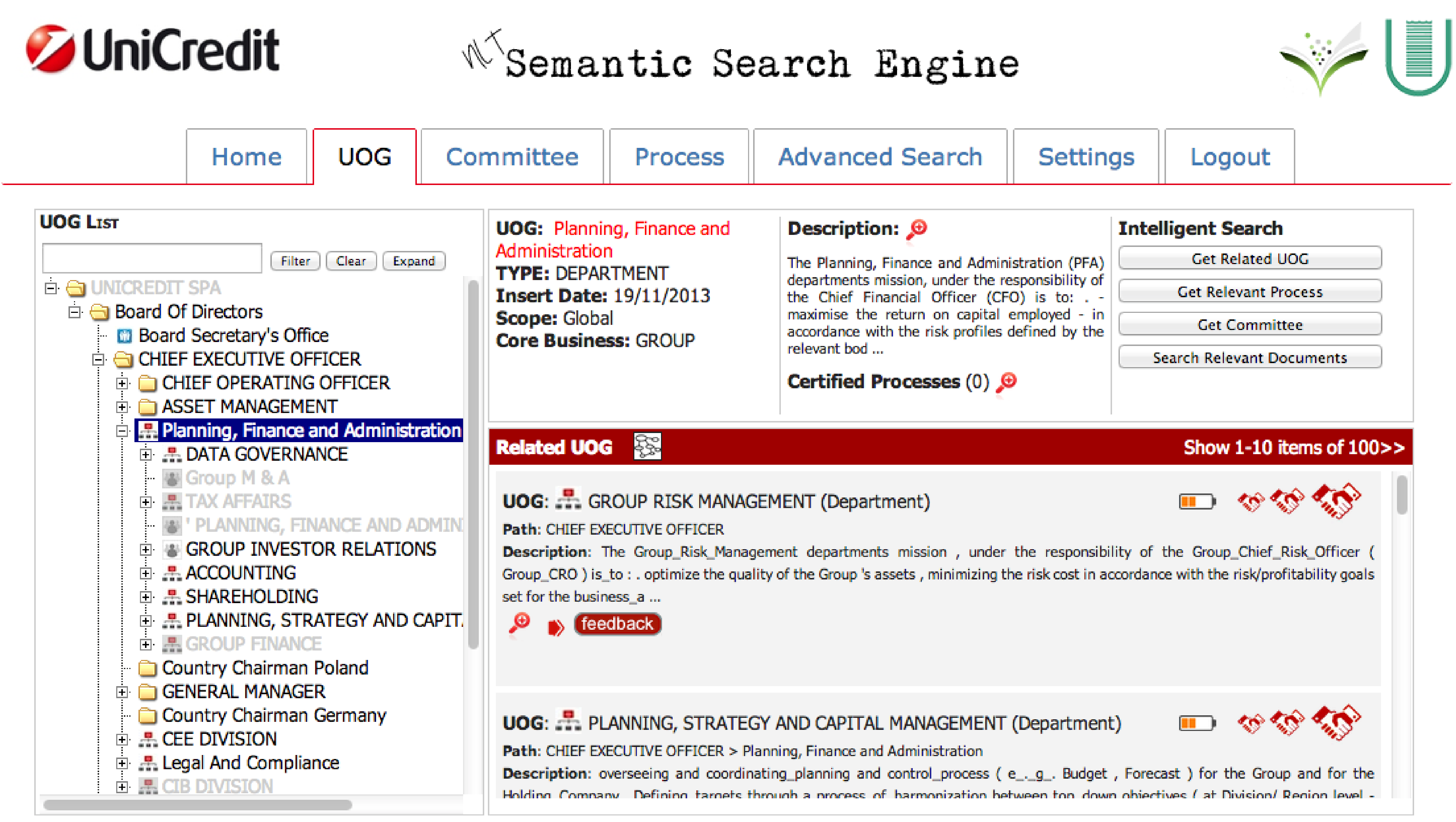
Dict



@biancofioremiky::u$$0.672
#sentenza::h$$0.755
#processoruby::h$$0.754
domondadori::h$$0.824
@generazionepdl::u$$0.757
#processomediaset::h$$0.836
@repubblica::u$$0.680
@iltweetpolitico::u$$0.689
#7anni::h$$0.771
@luigiamicone::u$$0.660
#incandidabilità::h
#tutticonsilvio::h$$0.763
#senzavergogna::h$$0.784
@fabriziaargano::u$$0.661
@alepaganotwit::u$$0.662
#nograzia::h$$0.760
#manifestazionepdl::h$$0.811
#leggeseverino::h$$0.760
@ddcnews::u$$0.664
@vincinoweb
@leggioggi_it::u$$0.668

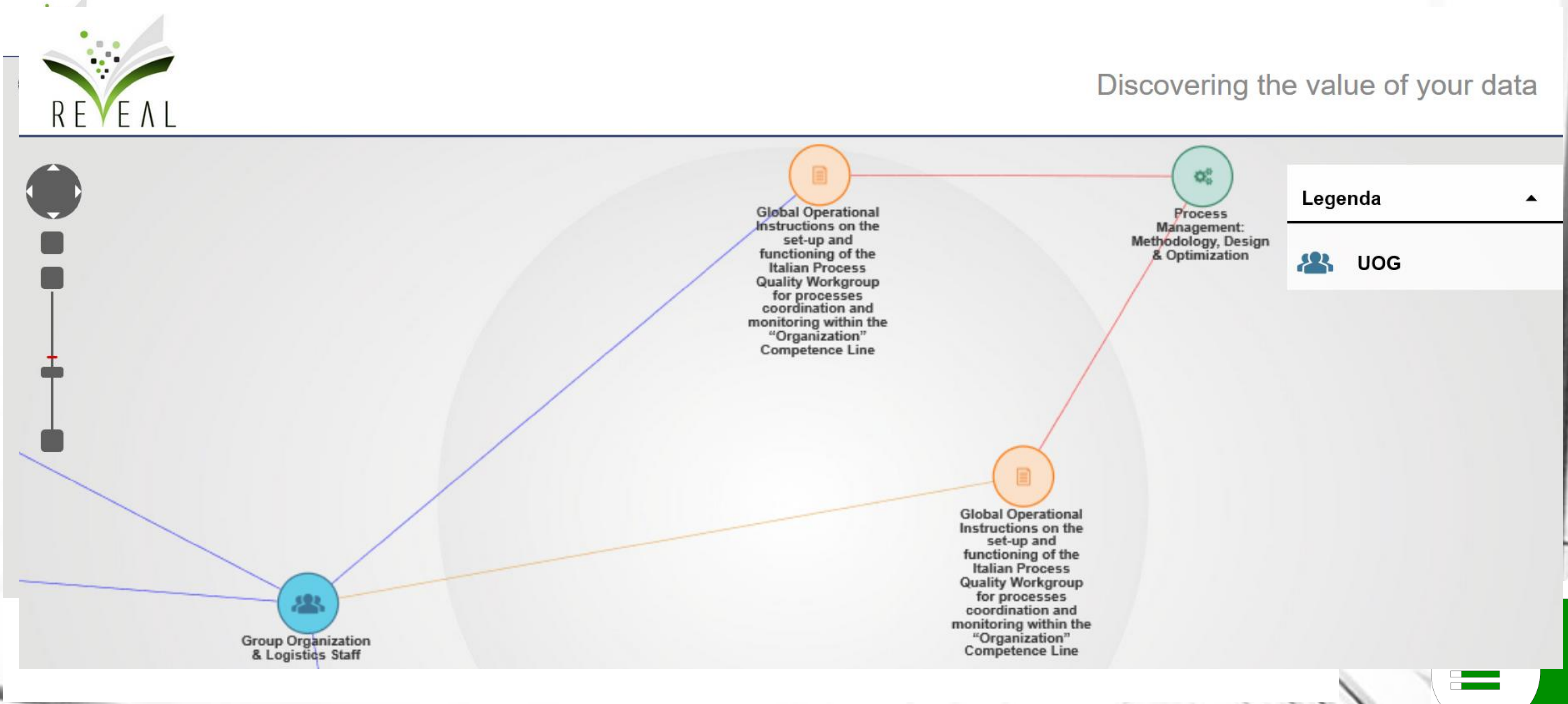
Search in Document Spaces

Semantic Search per le organizzazioni



Tor Vergata

Navigare il Knowledge Graph organizzativo



Tor Vergata

Crime Investigation

- Information and Relation Extraction da documentazione investigativa (ad es. sistemi informativi per la DNA)
  - Verbali di interrogatorio
  - Documenti giudiziari
  - Fonti aperte: Notizie stampa, Social media
- Riconoscimento e tracciamento di Entità, Eventi e Luoghi
- Semantic search
- Popolamento automatico e metadatazione di archivi distribuiti in supporto all’investigazione

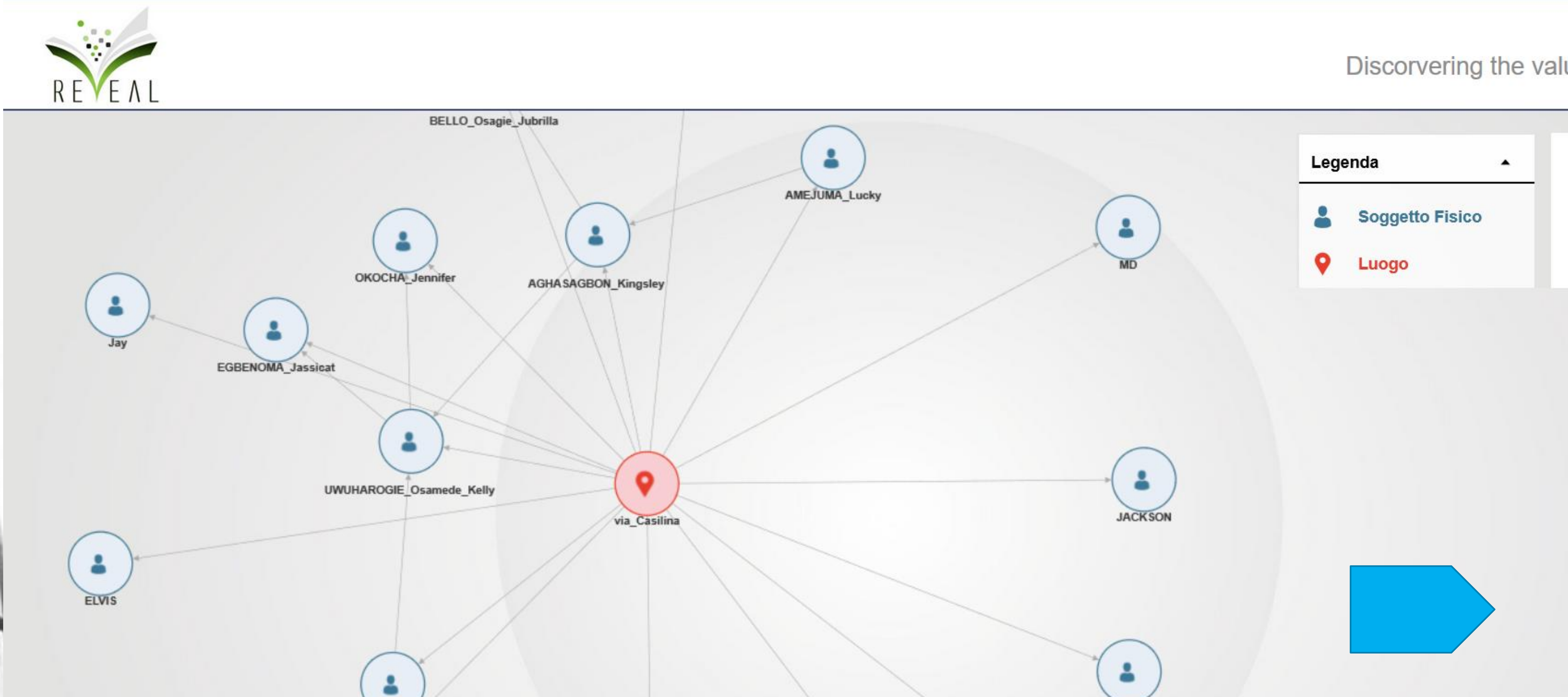
L " incontro veniva riscontrato da un servizio dinamico predisposto dal ROS CC Roma che alle ore 19,40 potevano notare AGHASAGBON_Kingsley in attesa nei pressi del lavaggio " Filosa Service " via_Giarratana nr . 50 , a bordo dell " autovettura Mercedes_190 targata CK_004_DB ; alle ore 19.45 , una persona riconosciuta in MASKAJ_Roland ed un altro uomo non identificato salivano a bordo del citato veicolo Mercedes_190 dell " AGHASAGBON , che si immetteva sulla via_Casilina e successivamente in via_Vermicino , fino ad arrivare al locale distributore " Q 8 " ; alle ore 20.15 , la vettura Mercedes_190 condotta da AGHASAGBON_Kingsley ripartiva dal distributore sino a giungere in via_Borgo_Fazio dove MASKAJ_Roland e l " altro uomo scendevano ; più tardi il veicolo Mercedes_190 , sempre condotto da AGHASAGBON_Kingsley veniva sottoposto a controllo da personale del Comando Compagnia Carabinieri di Frascati con esito negativo (Allegato 95) .

L"ascolto delle successive conversazioni permetteva di capire che mentre AGHASAGBON Kingsley incontrava il suo fornitore MASKAJ Roland, UWUHAROGIE Osamede Kelly attendeva altrove con il denaro destinato all"acquisto dello stupefacente

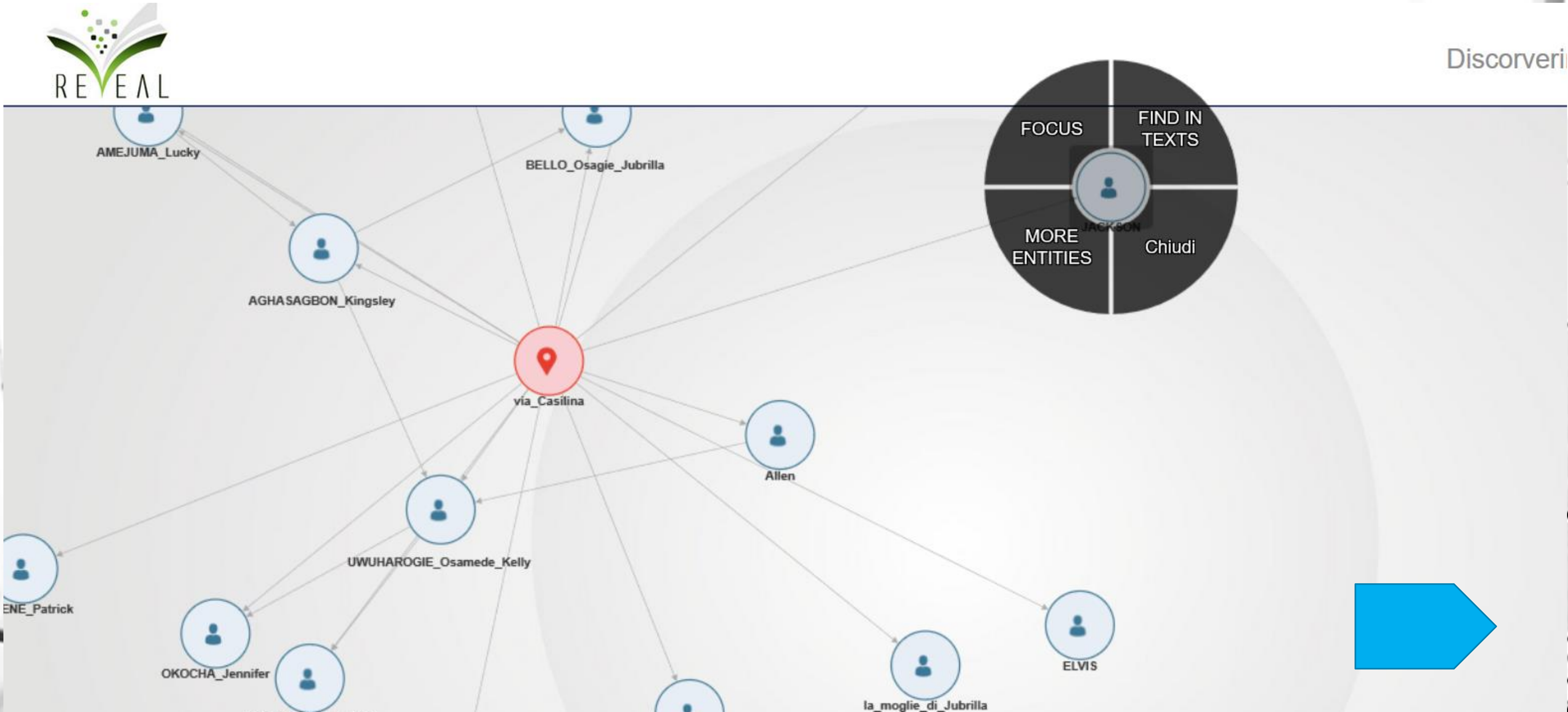
Retrieved Entities

NAME	TYPE	GRAPH
ROMA	FONTE_INFORMATIVA_DI_LUOGO	
AGHASAGBON KINGSLEY	FONTE_INFORMATIVA_DI_SOGGETTO_FISICO	
VIA GIARRATANA	FONTE_INFORMATIVA_DI_LUOGO	
MERCEDES BENZ E 190	AUTOMOBILE	
CK004DB	TARGA_AUTOMOBILISTICA	
MASKAJ ROLAND	FONTE_INFORMATIVA_DI_SOGGETTO_FISICO	
VIA CASILINA	FONTE_INFORMATIVA_DI_LUOGO	
VIA VERMICINO	FONTE_INFORMATIVA_DI_LUOGO	

Il grafo di una interrogazione e l'emergenza di comunità

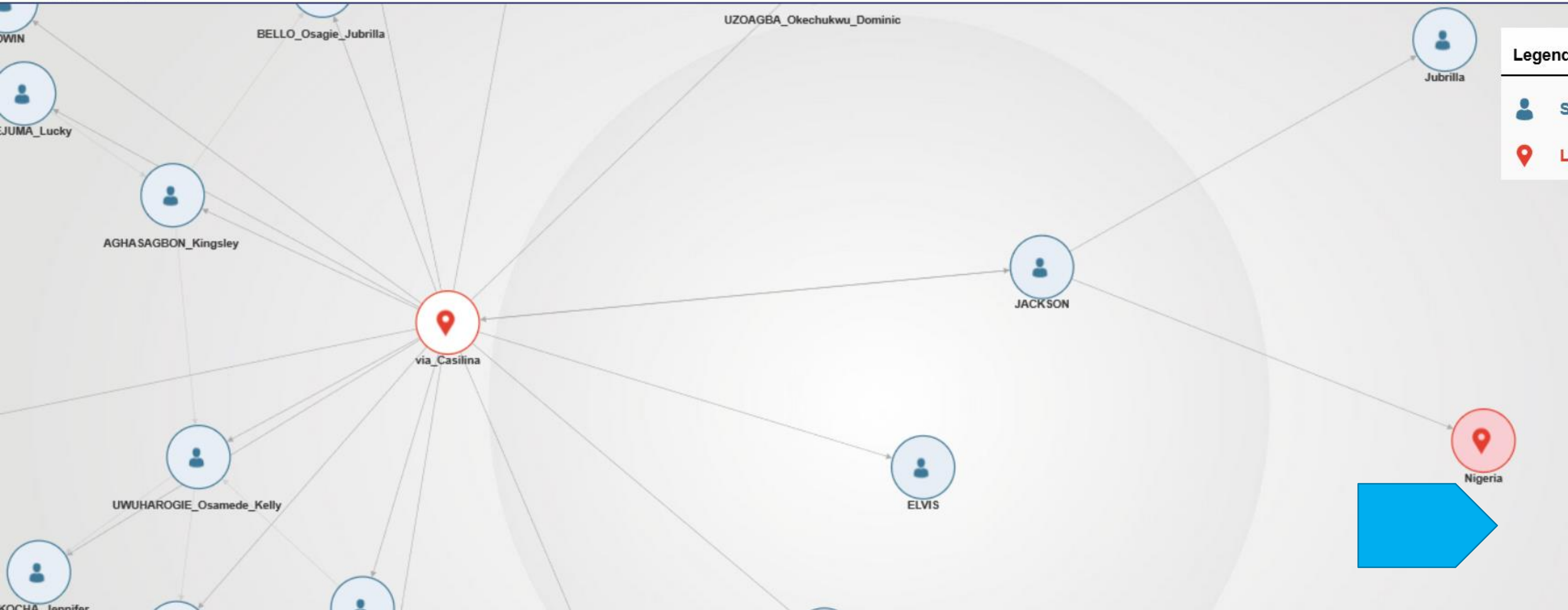


Funzionalità ai nodi



La esplorazione a partire da un nodo

Discorvering the value



Dai nodi di ritorno ai documenti

Search for:

Add Entity Constraint

Text	Type	
uwuharogie_osamede_kelly	persona	

Latest news

No news were found

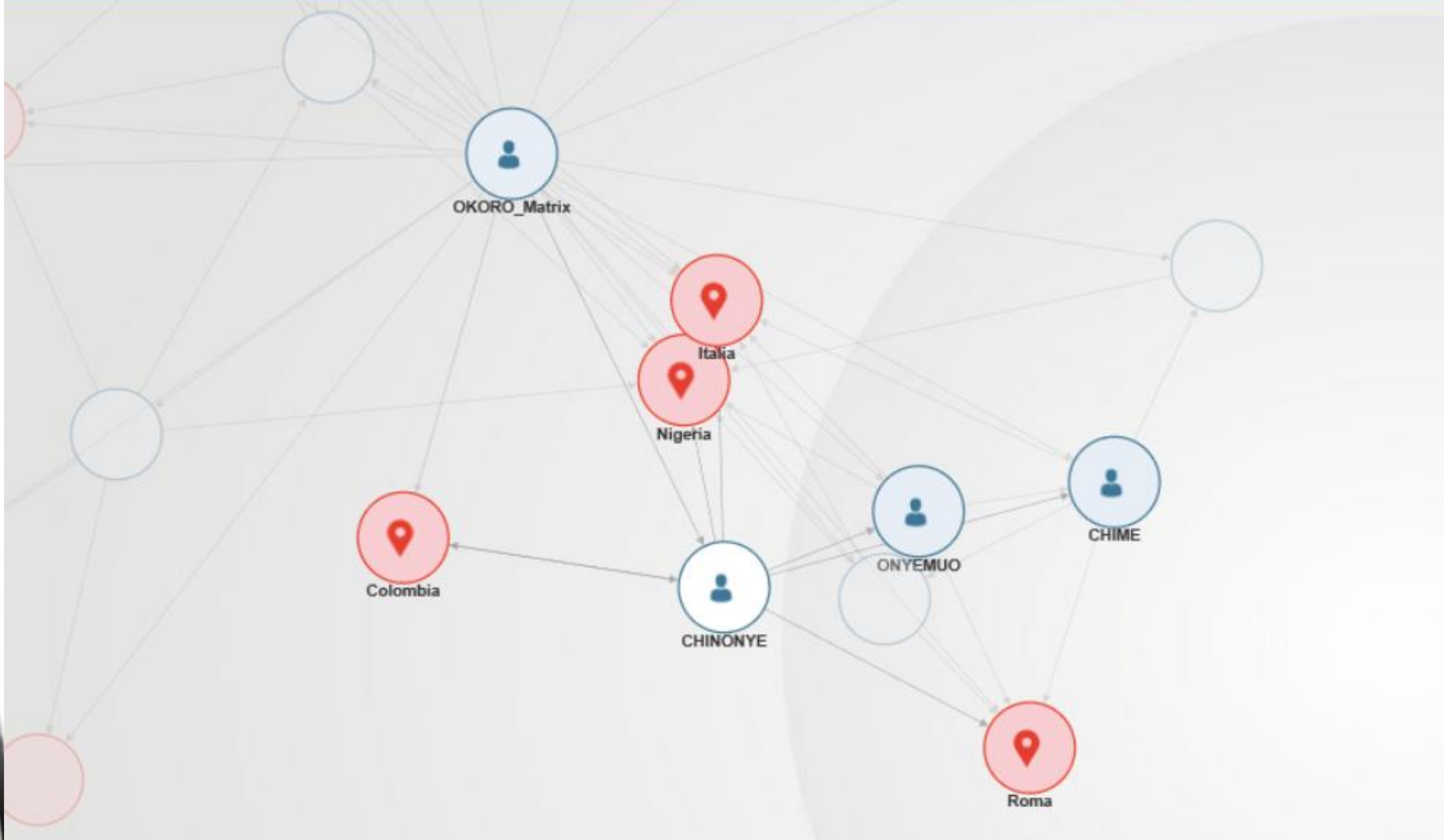
Document: Cult / 7.conclusioni.txt(13)

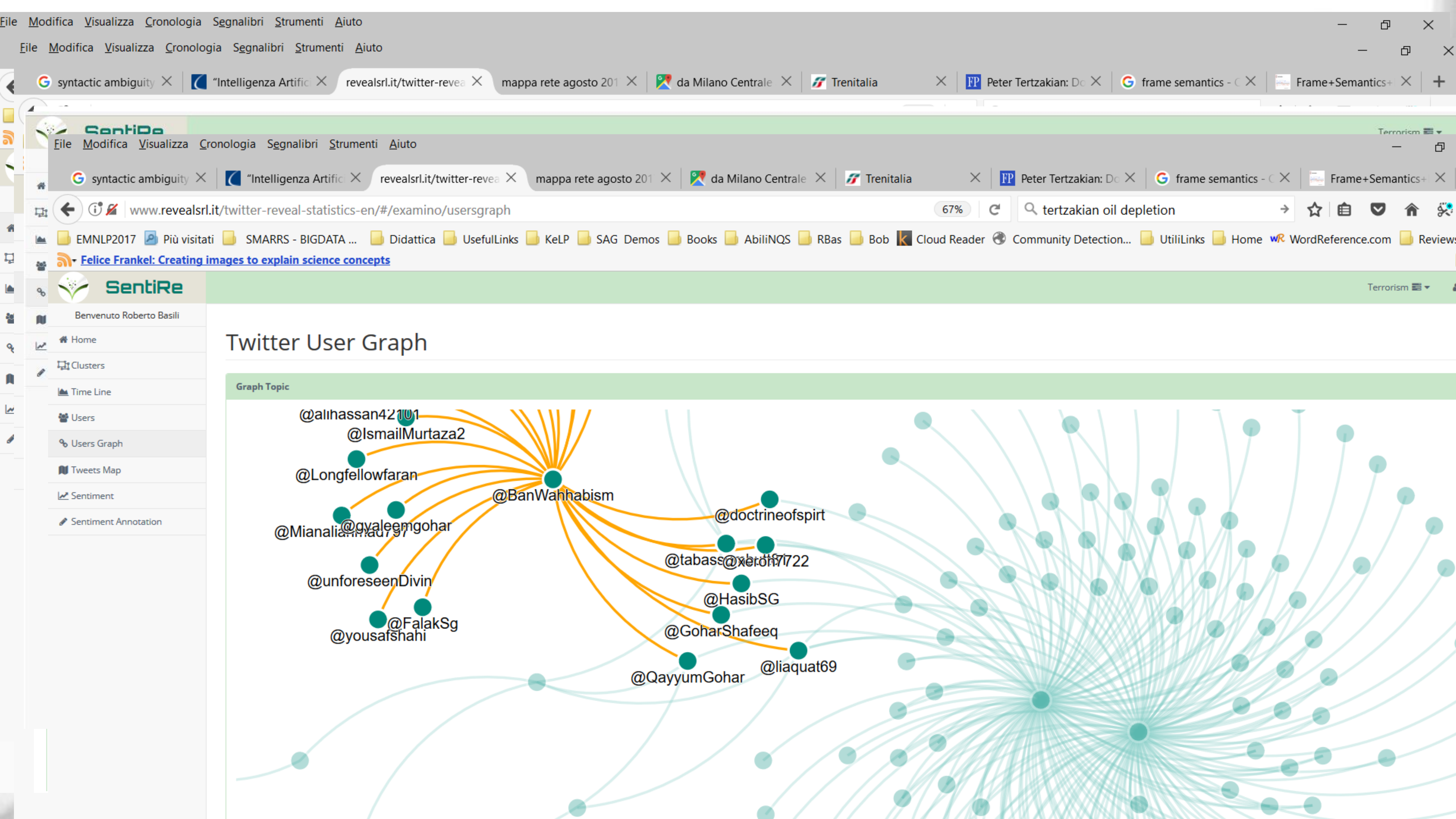
Best Paragraph

paragraph 179:

- UWUHAROGIE_OSAMEDE_Kelly è tra gli esponenti apicali di questa organizzazione dedita anche al traffico di sostanze stupefacenti , come peraltro segnalato dalla stessa collaboratrice . In_particolare UWUHAROGIE_Osamede_Kelly , dal call center denominato " Negozio di Kelly " , base operativa e centro di interessi dell " associazione , gestisce l " intero ciclo della commercializzazione dello stupefacente , approvvigionandosene prevalentemente da un gruppo di albanesi e distribuendolo , attraverso corrieri , alle ramificazioni territoriali dell " organizzazione , strumentali allo smercio e alla commercializzazione

Un esempio di informazione *implicita*




File Modifica Visualizza Cronologia Segnalibri Strumenti Aiuto
syntactic ambiguity
"Intelligenza Artific
revealsrl.it/twitter-revea
mappa rete agosto 201
da Milano Centrale
Trenitalia
Peter Tertzakian: Dc
frame semantics
Frame+Semantics
www.revealsrl.it/twitter-reveal-statistics-en/#/examino/usersgraph
67%
tertzakian oil depletion
EMNLP2017
Più visitati
SMARRS - BIGDATA ...
Didattica
UsefulLinks
KeLP
SAG Demos
Books
AbiliNQS
RBas
Bob
Cloud Reader
Community Detection...
UtiliLinks
Home
WordReference.com
Felice Frankel: Creating images to explain science concepts
SentiRe
Terrorism
Benvenuto Roberto Basili
Home
Clusters
Time Line
Users
Users Graph
Tweets Map
Sentiment
Sentiment Annotation
Twitter User Graph
Graph Topic
@IsmailMurtaza2
@Longfellowfaran
@BanWahhabism
@doctrineofspirt
@unforeseenDivin
@FalakSg
@yousafshahi
@HasibSG
@GoharShafeeq
@QayyumGohar
@liaquat69

Tendenze

- Integrazione di Linguaggio e Cognizione
  - Visione
  - Percezione: Grounding e Interazione Robotica
  - Pianificazione
- Deep Neural Learning in Semantic tasks
  - Reti ricorrenti e meccanismi di attenzione
  - Aumentare la sensiblità alle strutture linguistiche
  - Architetture end-to-end
- Social Computational Science
  - Demografia e Social Media
  - Studio del benessere

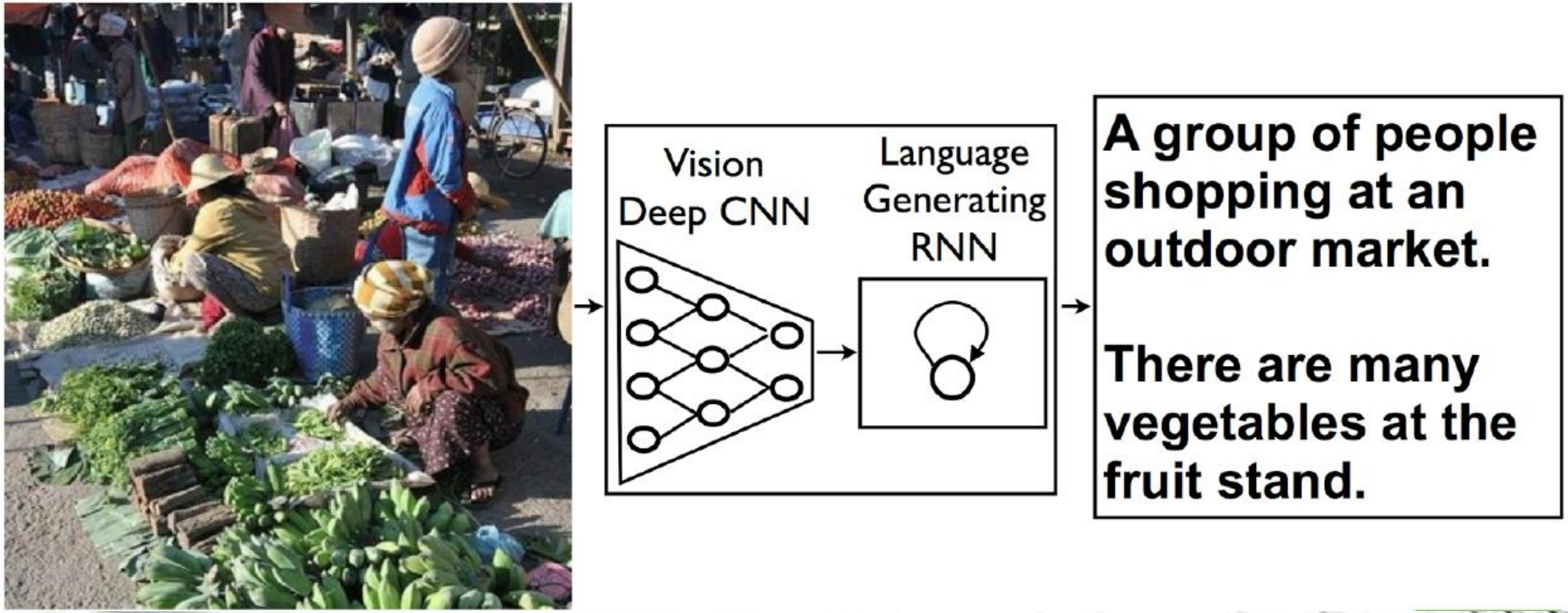
Un task *data-driven* molto complesso: *image captioning*

A caption is a brief **description** in natural language of an image which captures the objects, the relations between them and the actions performed by them, and is grammatically and synctactically correct.

"A chef preparing food inside of a kitchen near a window."

Automatic image captioning (Vinyals et al., 2014)

In the recent years, ML has offered the chance to build an *integrated solution* for image captioning, that performs these subtasks simultaneously by using **neural networks**.



… una rete neurale addestrata per il *captioning* in Italiano (Masotti et al., 2017)

Uno scuolabus giallo parcheggiato sul lato della strada.

Un uomo che cavalca un cavallo su una strada cittadina.

Un segnale di stop che si siede su un angolo di strada.

References

- AI & Robotics. «Robot Futures», Ilah Reza Nourbakhsh, MIT Press, 2013
- NLP & ML:
  - «Statistical Methods for Speech Recognition», F. Jelinek, MIT Press, 1998
  - «Speech and Language Processing", D. Jurafsky and J. H .Martin, Prentice-Hall, 2009.
  - "Foundations of Statistical Natural Language Processing, Manning & Schtze, MIT Press 2001.
- Sitografia:
  - SAG, Univ. Roma Tor Vergata: http://sag.art.uniroma2.it/
  - Reveal s.r.l.: http://www.revealsrl.it/