

Strumenti per l'Estrazione di Informazione dai Testi

Rachele Sprugnoli – rachele.sprugnoli@unicatt.it

Centro Interdisciplinare di Ricerche per la Computerizzazione
dei Segni dell'Espressione (CIRCSE)



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

CHI SONO



University of Pisa	CELCT	Fondazione Bruno Kessler	University of Trento	Catholic University
Bachelor's and Master's Degree in Humanities Computing	Center for the Evaluation of Language and Communication Technologies	Digital Humanities Research Group	PhD in Information Technology	Interdisciplinary Research Center for the Computerization of the Signs of Expression



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme – Grant Agreement No 769994

COSA FAREMO

6 MARZO: Introduzione teorica e uso di una pipeline (demo online)

13 MARZO: Uso di una pipeline per l'italiano + focus su NER

20 MARZO: Estrazione di temi e parole chiave

27 MARZO: Estrazione del sentiment

ESTRAZIONE DI INFORMAZIONE DA TESTI

Processo che consente di ottenere dati strutturati da un documento (DIGITALE) in linguaggio naturale non strutturato

Si estraggono solo le informazioni di interesse per un certo scopo: entità, relazioni, eventi, date, temi...

Dati NON strutturati:

- articoli di giornale, post sui social media, articoli scientifici, testi letterari, fonti storiche...

Dati strutturati:

- tabelle, template, schemi, database

ESEMPIO

ROMA - Simona Halep è la regina di Roma. La tennista romena al terzo tentativo si è aggiudicata la 77esima edizione degli Internazionali d'Italia femminili. In finale Halep, numero 2 del ranking e prima favorita del seeding, ha superato Karolina Pliskova, n.4 Wta e seconda testa di serie, nonché campionessa in carica. La ceca è stata costretta al ritiro per infortunio sul punteggio di 6-0 2-1 per la 28enne di Costanza dopo appena 32 minuti di gioco.



- Persons: Simona Halep, Karolina Pliskova
- Locations: Roma, Costanza
- Events: 77esima edizione degli Internazionali d'Italia femminili

APPLICAZIONI

PUBBLICA AMMINISTRAZIONE: identificare trend nell'attività della PA, semplificare l'accesso alle informazioni

BIBLIOTECHE: riconoscere autori/riferimenti bibliografici, individuare articoli pertinenti

MARKETING: individuare dove un prodotto è menzionato, identificare l'attitudine dei clienti verso un brand

SANITÀ: individuare sintomi e terapie dalle cartelle cliniche

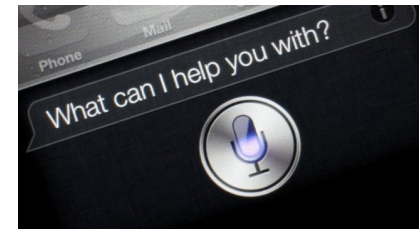
STORIA: estrarre eventi dalle fonti, individuare fonti su argomenti simili

MA...

Il computer, di per sé, **NON** conosce il linguaggio naturale!

Il **Trattamento Automatico del Linguaggio** (TAL) ha lo scopo di dotare il computer di conoscenze linguistiche, di creare macchine che capiscano (e addirittura riproducano) il linguaggio naturale, di sviluppare programmi che assistano l'essere umano in compiti (*task*) linguistici:

- riconoscimento automatico del parlato
- sintesi automatica della voce
- traduzione automatica
- analisi automatica del sentimento



PERCHÉ È UNA SFIDA

1. Ambiguità grammaticale

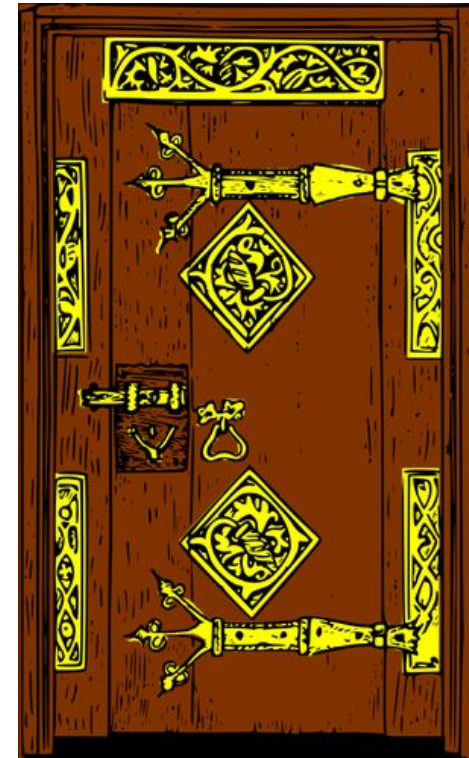
PAROLA	CATEGORIA GRAMMATICALE
C'	AVVERBIO/PRONOME
era	VERBO/NOME
una	ARTICOLO/PRONOME/NUMERALE
volta	NOME/VERBO (voltare)/VERBO (volgere)
un	ARTICOLO/NUMERALE
pezzo	NOME
di	PREPOSIZIONE
legno	NOME/VERBO

PERCHÉ È UNA SFIDA

2. Ambiguità sintattica: «una vecchia porta la sbarra»



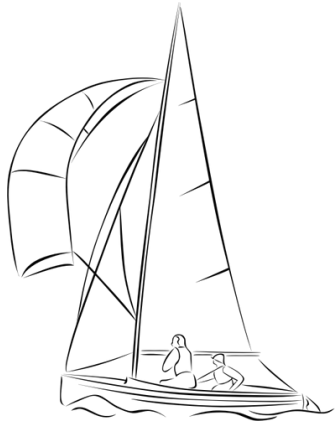
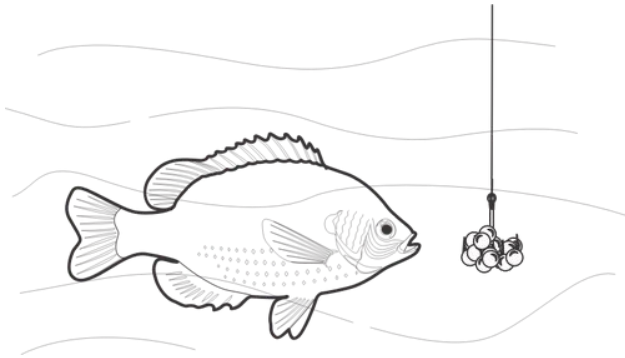
«una **vecchia** porta la sbarra»



«una vecchia **porta** la sbarra (la strada)»

PERCHÉ È UNA SFIDA

3. Ambiguità semantica: «*amo*» / «*navigare*»



PERCHÉ È UNA SFIDA

4. La lingua cambia

- Lingue classiche/storiche:

*Ahi quanto a dir qual era è cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinova la paura!*



- Lingue non-standard:

[#SanremoFunky](#) con [@elodie](#) e qualche considerazione sulla prima serata di [#Sanremo2020](#) 🎵 che sta per partire

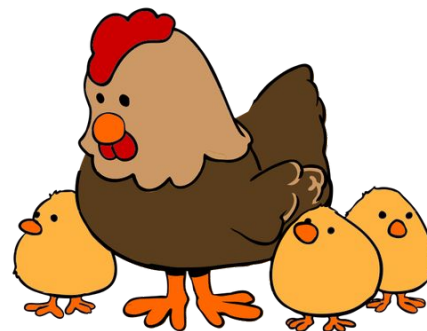
- Neologismi: *petaloso* / *Brexit*

PERCHÉ È UNA SFIDA

5. Espressioni multi-parola, ovvero «2 +2 non fa sempre 4»

Il loro significato non corrisponde alla combinazione lessicale delle parole che li compongono

- espressioni metaforiche: «*parlare dietro le spalle*»
- proverbi: «*si salvi chi può*»
- espressioni idiomatiche: «*conosco i miei polli*»



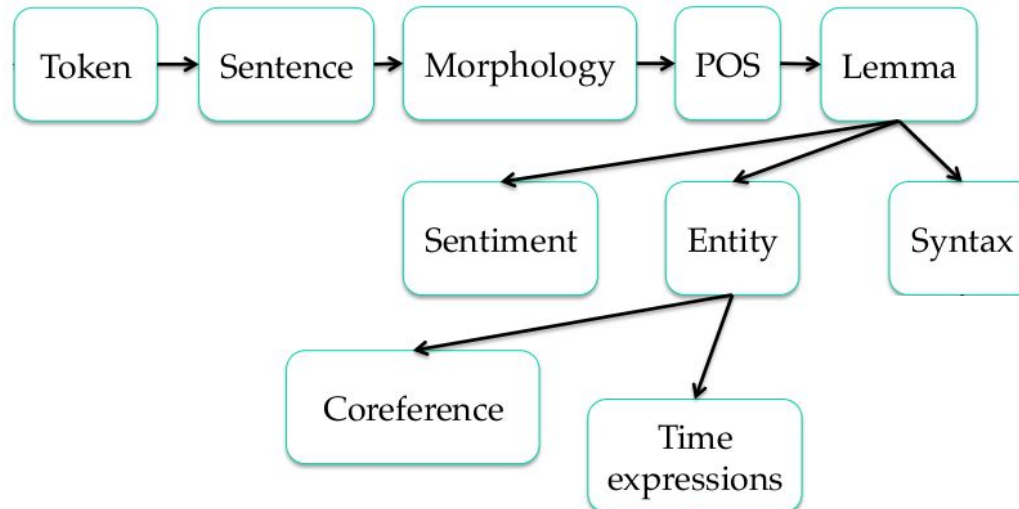
PERCHÉ È UNA SFIDA

6. Servono informazioni di contesto o di conoscenza del mondo
«Elsa e Anna sono sorelle»



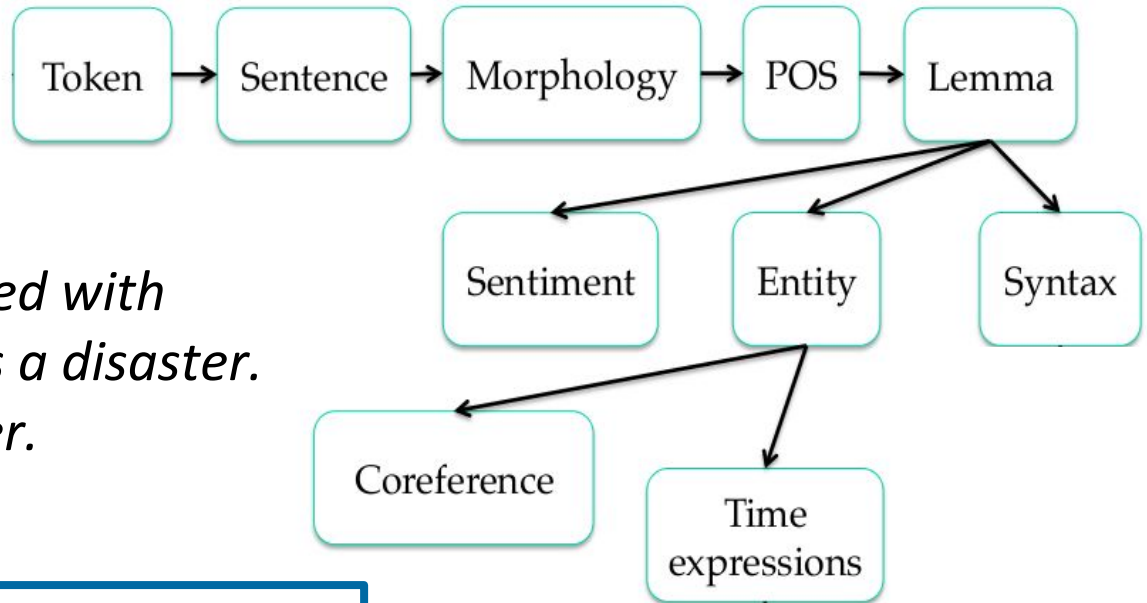
COME ANALIZZARE IL LINGUAGGIO

- Struttura a **PIPELINE**: catena i cui moduli descrivono ognuno un diverso livello di analisi linguistica e dove l'output di un modulo diventa l'input per il modulo successivo. Esempio:



Le analisi presentate nelle prossime slide sono l'output della pipeline di Stanford CoreNLP

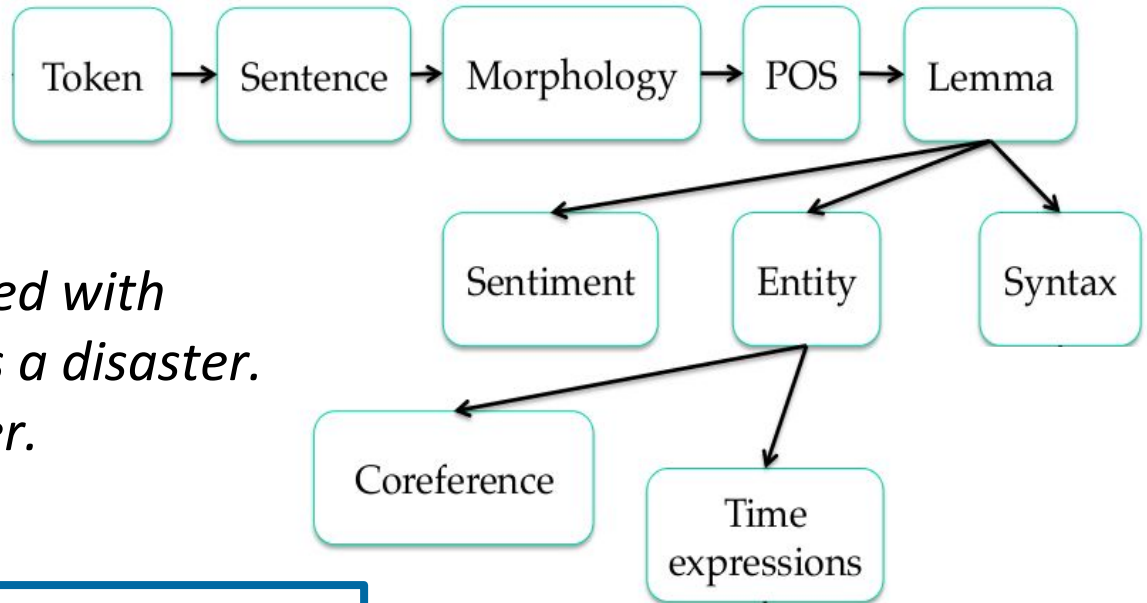
- demo online: <http://corenlp.run/>



When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
 Trump, 2016-08-05

TOKEN - SENTENCE - PART OF SPEECH

	WRB	PRP	VBP	WP	VBD	IN	JJ	NNP	NN	,	PRP	VBD	DT	NN	.
1	When	you	see	what	happened	with	crooked	Hillary	today	,	it	was	a	disaster	.
2	DT	NN	.												
	A	disaster	.												
3	PRP	VBD	DT	NN	.										
	She	had	a	disaster	.										



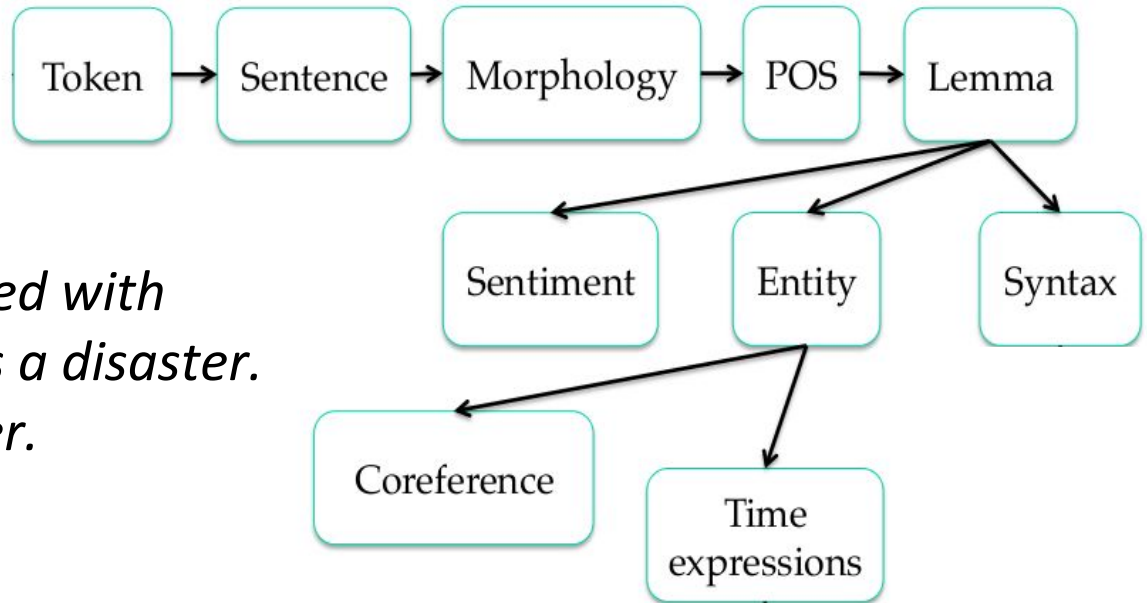
When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
Trump, 2016-08-05

TOKEN - SENTENCE - PART OF SPEECH

C'era una volta un pezzo di legno.

C'era | una | volta | un | pezzo | di | legno.

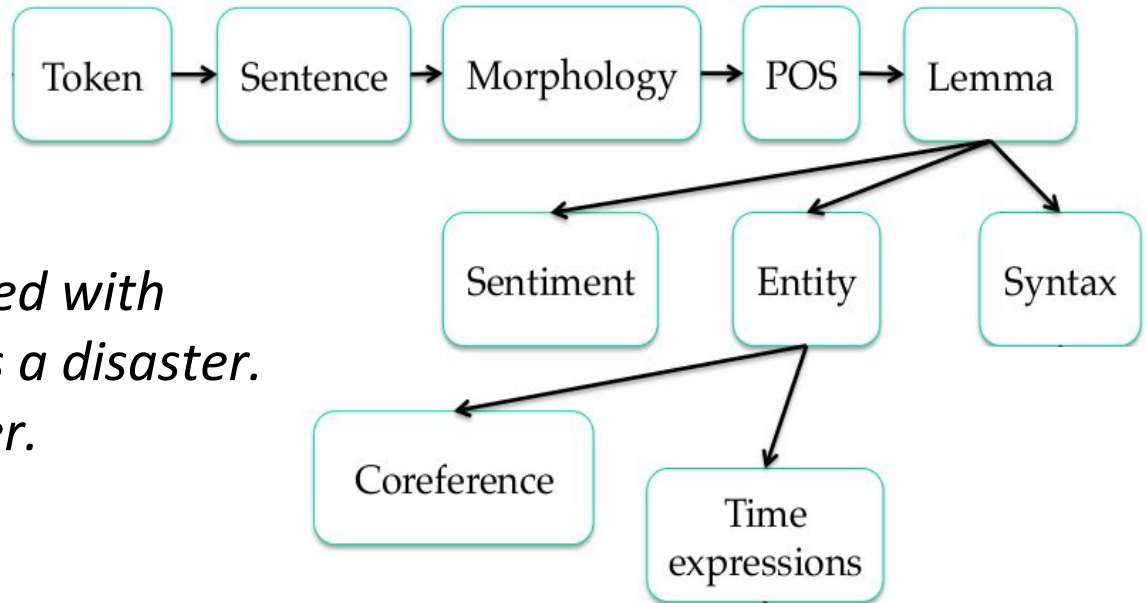
C' | era | una | volta | un | pezzo | di | legno | .



When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
 Trump, 2016-08-05

MORPHOLOGY

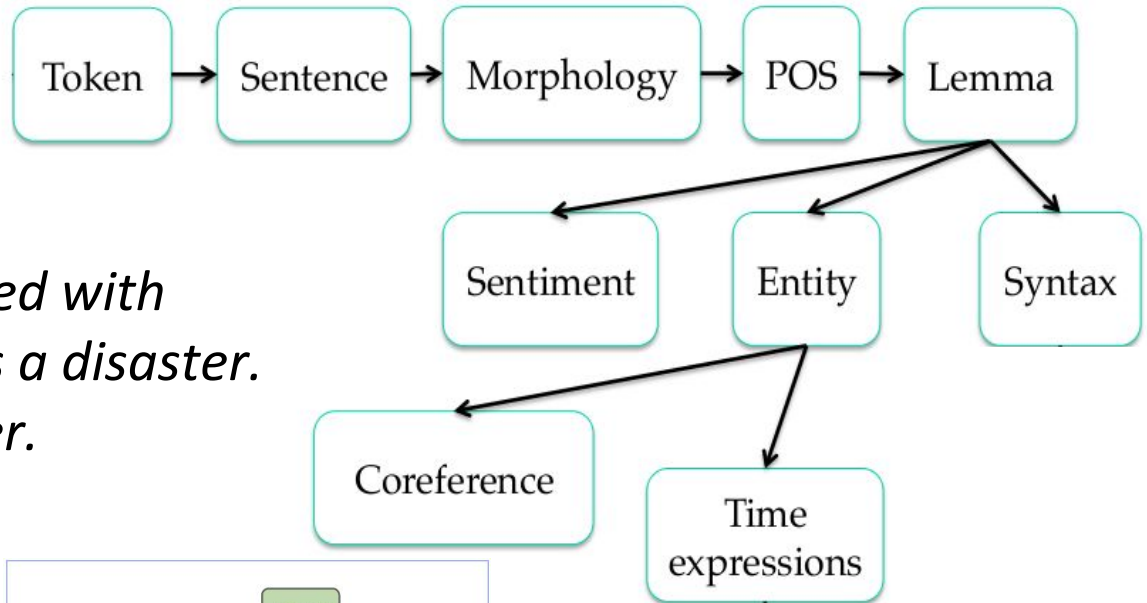
when+conj	you+pron	see+v+indic+pres+no3sing	what+adj+zero	happen+v+indic+past	with+prep	crooked+adj+zero	NULL	today+adv	NULL
When	you	see	what	happened	with	crooked	Hillary	today	,
it+pron	be+v+indic+past	a+art	disaster+n+sing	disaster	disaster	disaster	.	.	.
it	was	a	disaster
a+art	disaster+n+sing	disaster
A	disaster
she+pron	have+v+indic+past	a+art	disaster+n+sing	disaster	disaster	disaster	.	.	.
She	had	a	disaster



When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
Trump, 2016-08-05

LEMMA

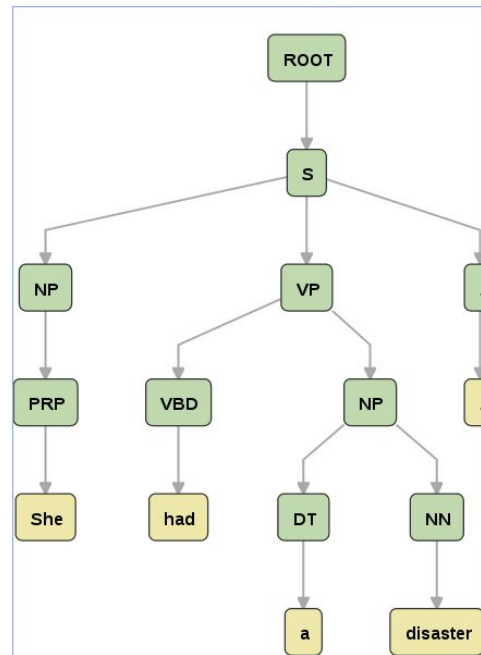
- 1 when you see what happen with crooked Hillary today , it be a disaster .
- 2 a disaster .
- 3 she have a disaster .

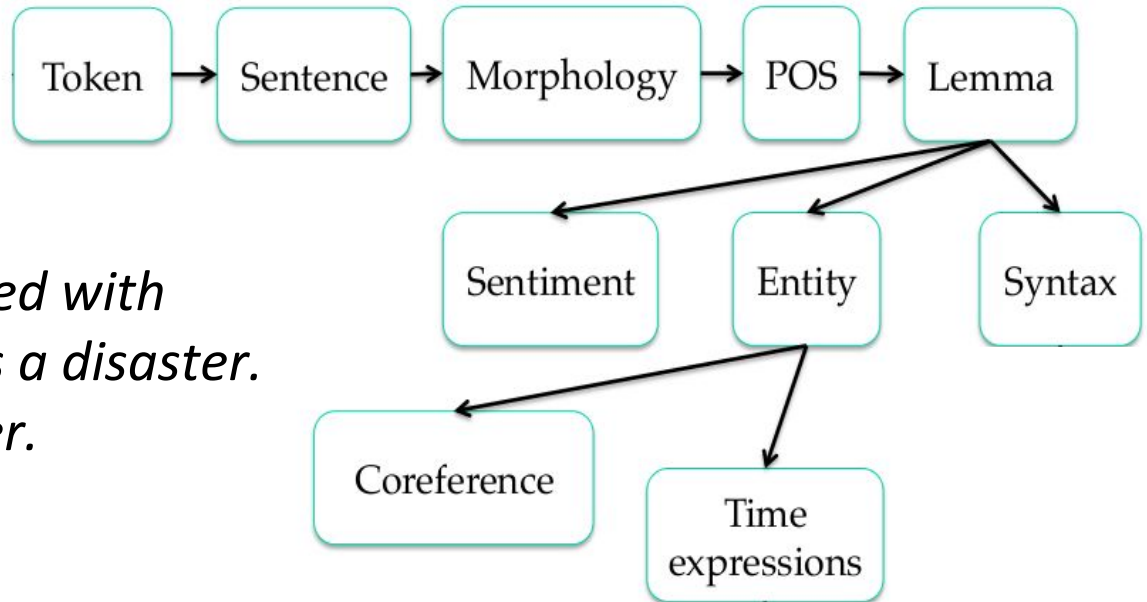


When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
Trump, 2016-08-05

SYNTAX / PARSING

- a costituenti

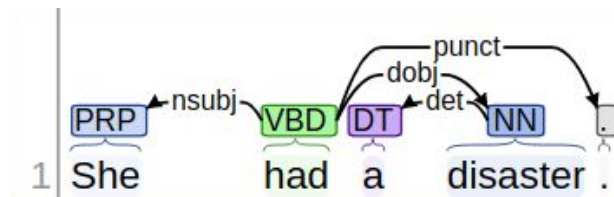


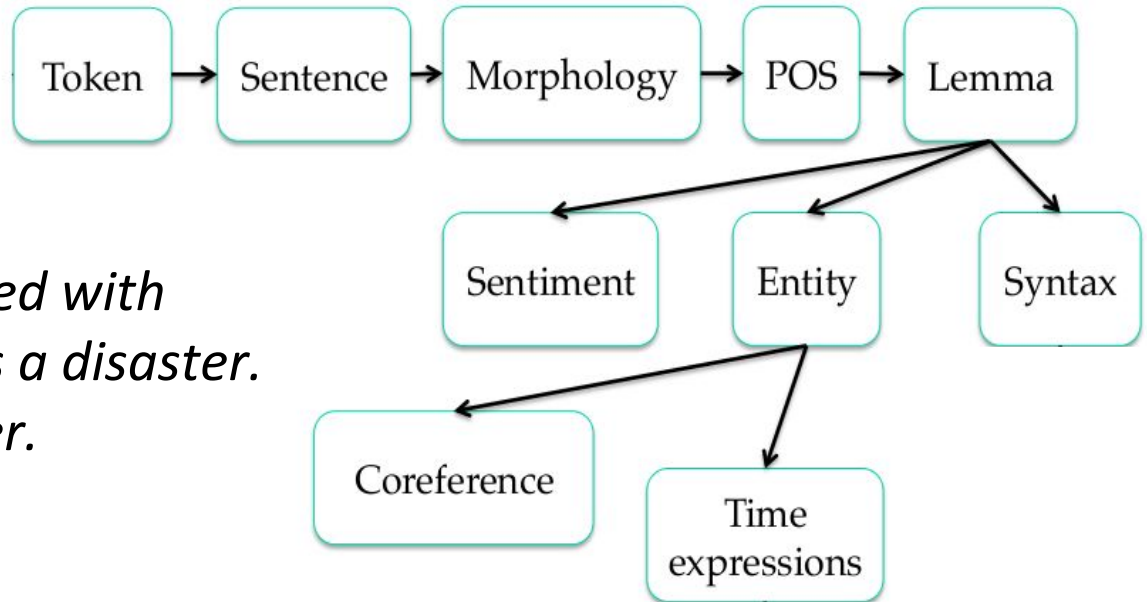


When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
Trump, 2016-08-05

SYNTAX / PARSING

- a dipendenze

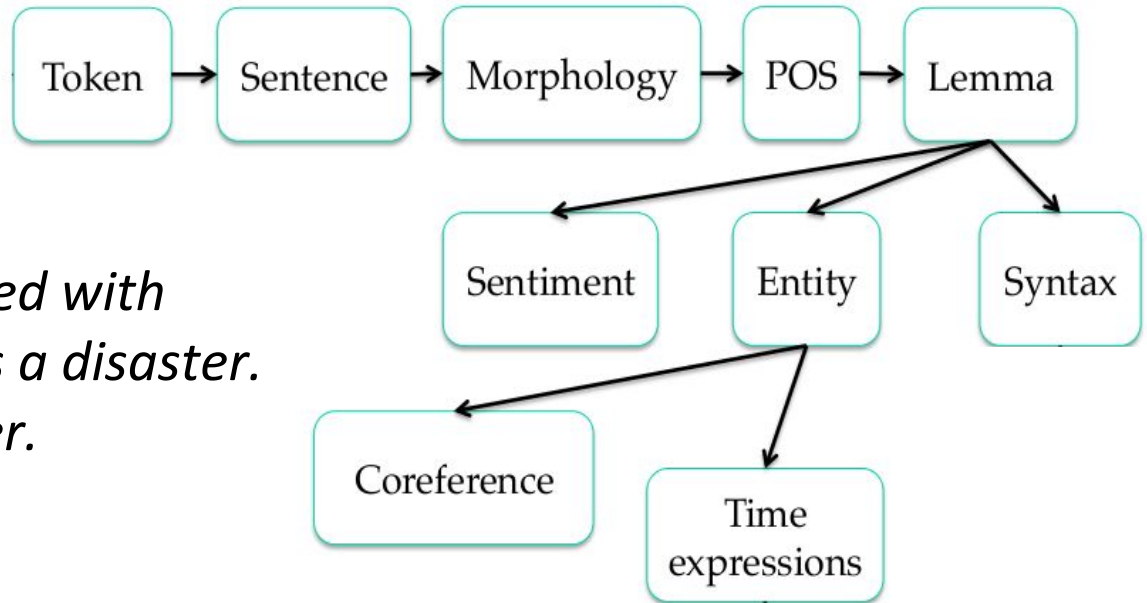




When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
Trump, 2016-08-05

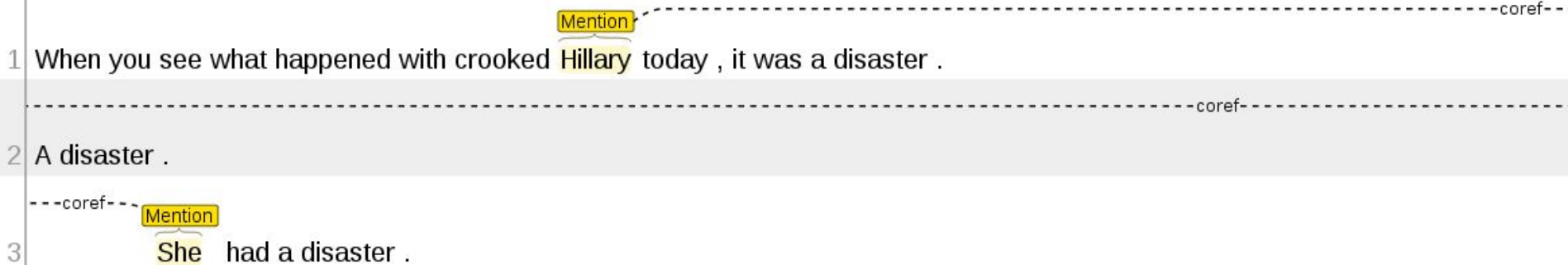
ENTITY

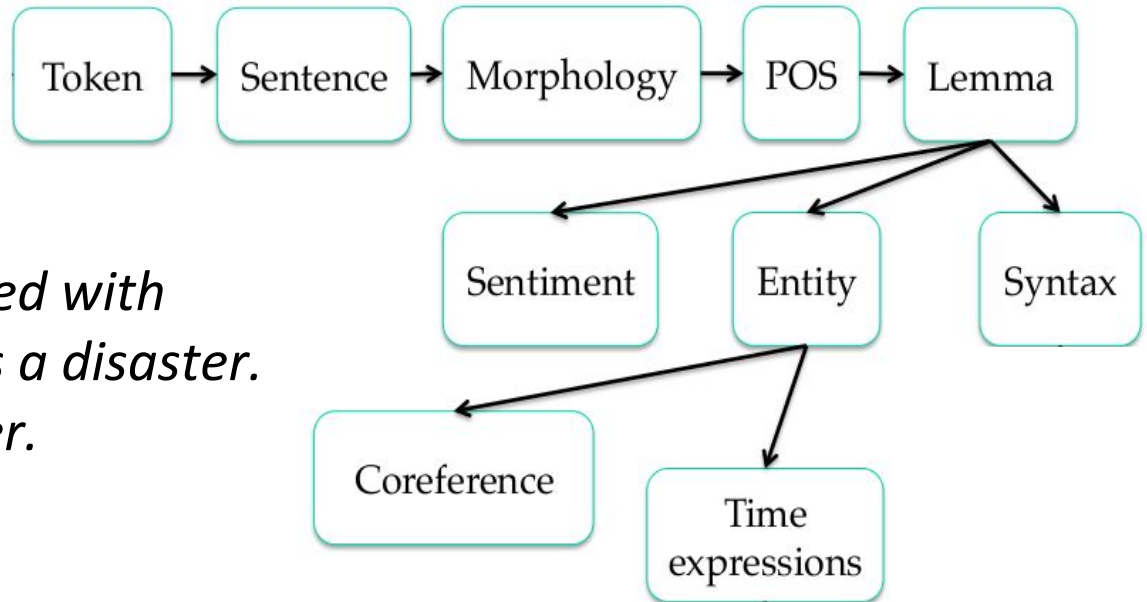
- 1 When you see what happened with crooked PER Hillary today , it was a disaster .
- 2 A disaster .
- 3 She had a disaster .



When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
Trump, 2016-08-05

COREFERENCE



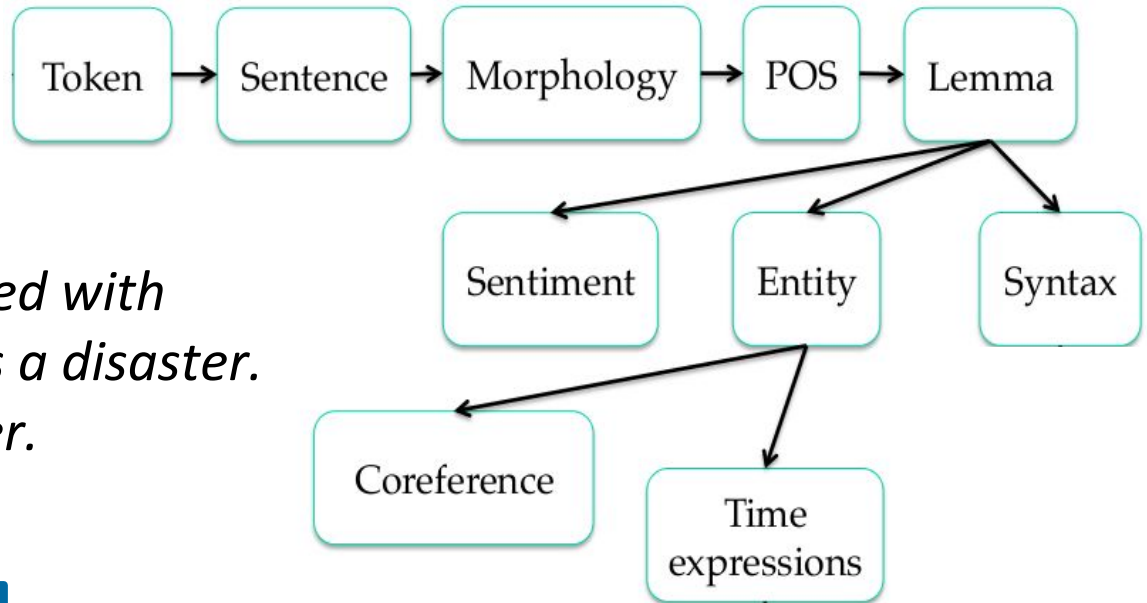


When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
Trump, 2016-08-05

TIME EXPRESSIONS

2016-08-05

- | | |
|---|--|
| 1 | When you see what happened with crooked Hillary <u>today</u> , it was a disaster . |
| 2 | A disaster . |
| 3 | She had a disaster . |



When you see what happened with crooked Hillary today, it was a disaster. A disaster. She had a disaster.
Trump, 2016-08-05

SENTIMENT

		NEGATIVE
1	When you see what happened with crooked Hillary today , it was a disaster .	
2	A disaster .	VERY NEGATIVE
3	She had a disaster .	NEGATIVE

COME SI SVILUPPA UN MODULO TAL

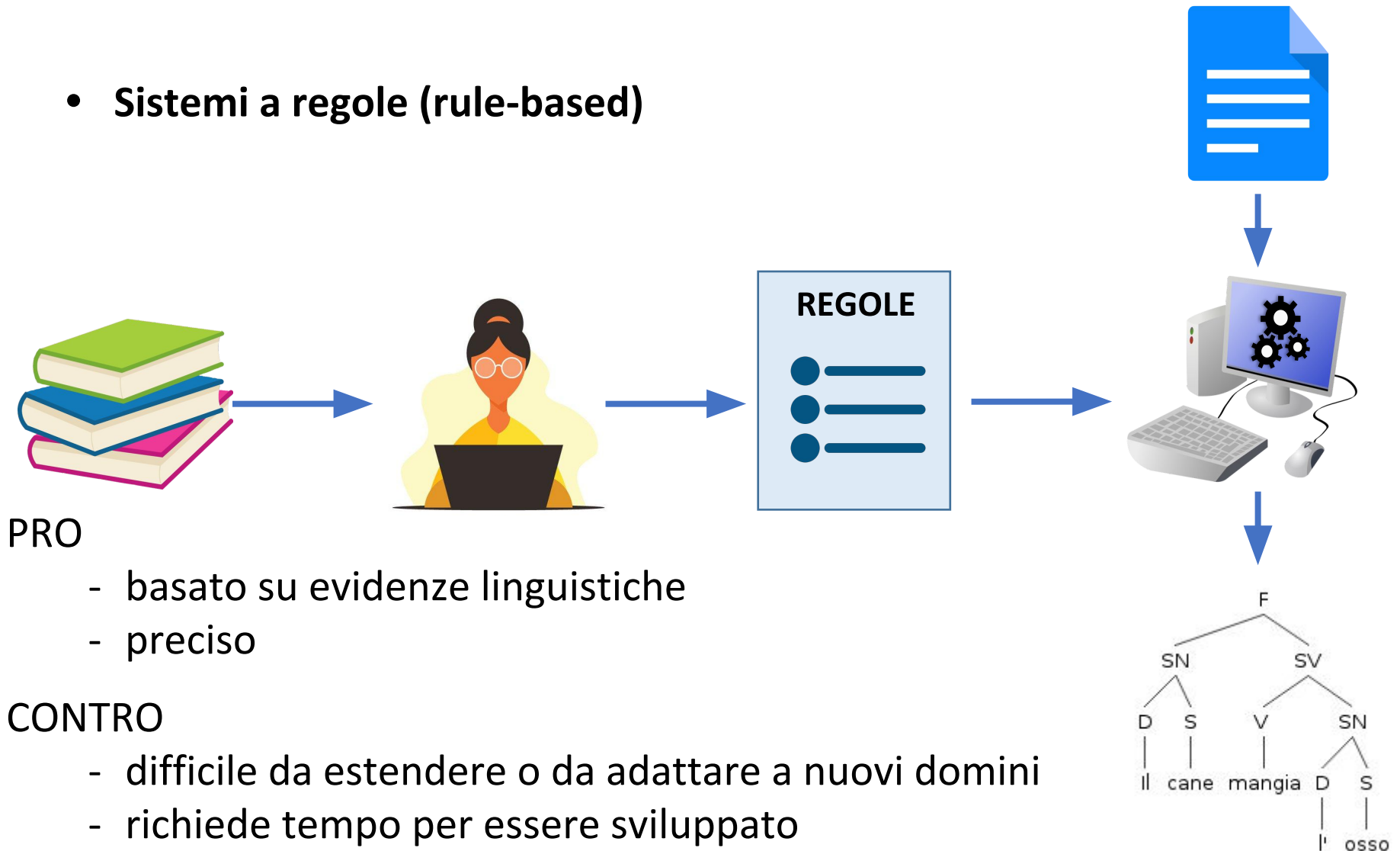
- **LOOKUP LIST**

- Sistema che riconosce solo le parole memorizzate nei suoi elenchi detti “gazetteers”
- Vantaggi: semplice, veloce, facile da utilizzare
- Svantaggi: la raccolta e il mantenimento degli elenchi richiede tempo, gli elenchi non gestiscono tutte le possibili varianti delle parole e non possono risolvere l'ambiguità, nessun tipo di inferenza

LISTA_VALUTE	LISTA_CITTÀ
Euro, dollaro, dollari, sterlina, sterline, \$, €...	<u>http://download.geonames.org/export/dump/</u>

COME SI SVILUPPA UN MODULO TAL

- Sistemi a regole (rule-based)



COME SI SVILUPPA UN MODULO TAL

- **Sistemi a regole (rule-based)**
- Esempio: Part-of-Speech tagging:
 - 1) assegnazione ad ogni parola di tutti i possibili PoS usando un dizionario

		NOUN
VERB	ART	VERB
«paghiamo	il	conto»

- 2) applicazione delle regole per rimuovere etichette ambigue
 - «rimuovere VERB se in alternativa con NOUN e preceduto da ART»

		NOUN
VERB	ART	VERB
«paghiamo	il	conto»

COME SI SVILUPPA UN MODULO TAL

- Sistemi a regole (rule-based)

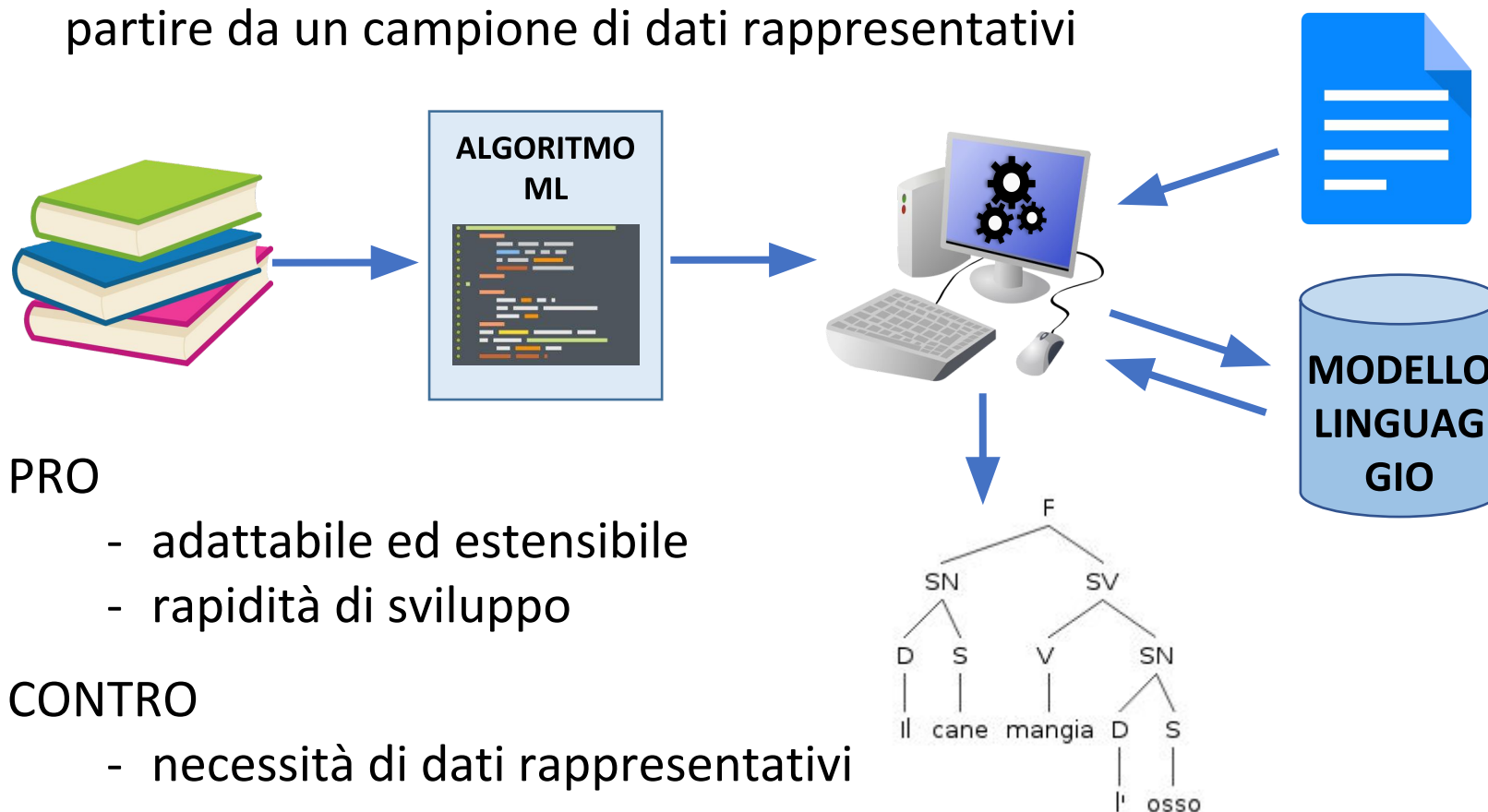
Named Entity Recognition without Gazetteers,
Mikheev et al. 1999

Context Rule	Assign	Example
Xxxx+ is? a? JJ* PROF	PERS	Yuri Gromov, a former director
Xxxx+ is? a? JJ* REL	PERS	John White is beloved brother
Xxxx+ himself	PERS	White himself
Xxxx+, DD+, shares in Xxxx+	PERS	White, 33,
PROF of/at/with Xxxx+	ORG	shares in Trinity Motors
Xxxx+ area	LOC	director of Trinity Motors
		Beribidjan area

COME SI SVILUPPA UN MODULO TAL

- **Sistemi di apprendimento automatico – MACHINE LEARNING (ML)**

- algoritmi che permettono al computer di imparare a svolgere un task a partire da un campione di dati rappresentativi



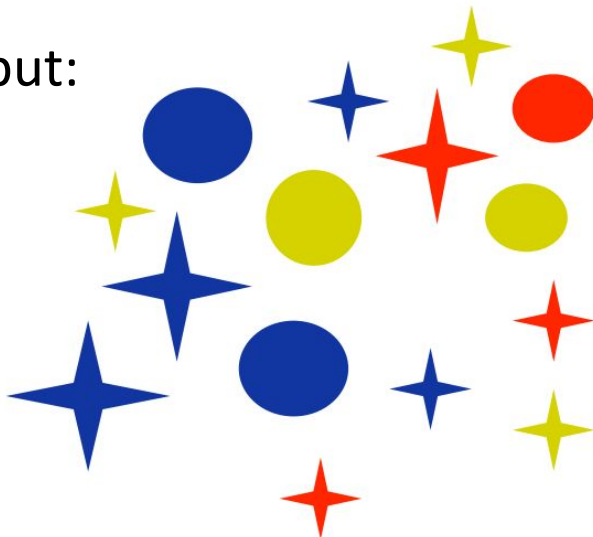
COME SI SVILUPPA UN MODULO TAL

- **Sistemi di apprendimento automatico – MACHINE LEARNING (ML)**
 - **3 tipi principali di algoritmi di ML**
1. **NON SUPERVISIONATI:** non necessitano di un corpus annotato a mano per creare il modello
 2. **SUPERVISIONATI:** utilizzano un corpus annotato a mano per la creazione dei modelli
 3. **SEMI-SUPERVISIONATI:** combinano informazioni derivanti sia da corpora annotati che da dati non annotati

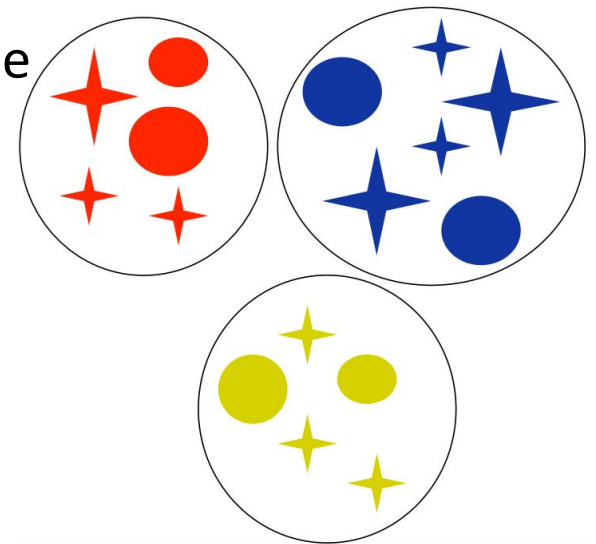
COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML NON SUPERVISIONATO, esempio
 - CLUSTERING: raggruppamento dell'input in base a una qualche relazione di similitudine tra i dati

Input:



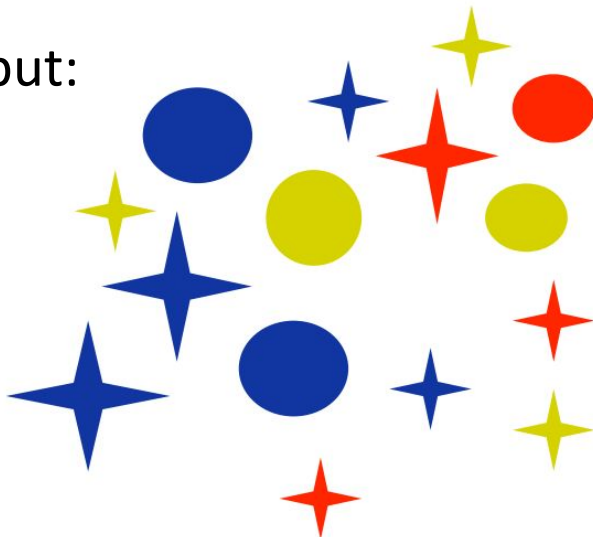
Output in base
al colore:



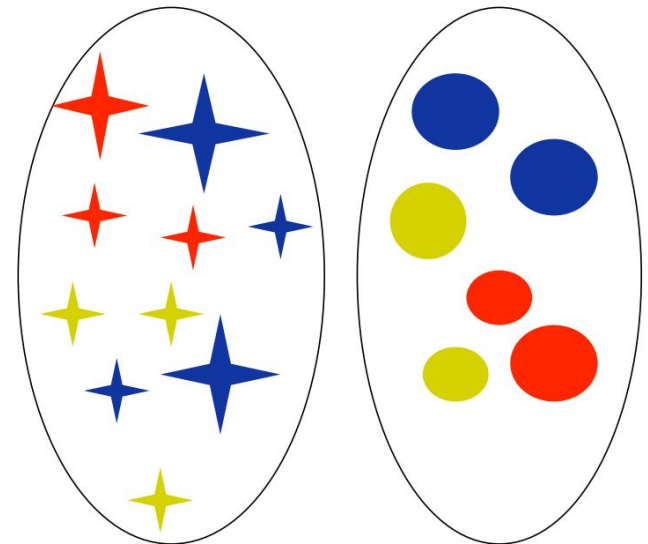
COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML NON SUPERVISIONATO, esempio
 - CLUSTERING: raggruppamento dell'input in base a una qualche relazione di similitudine tra i dati

Input:

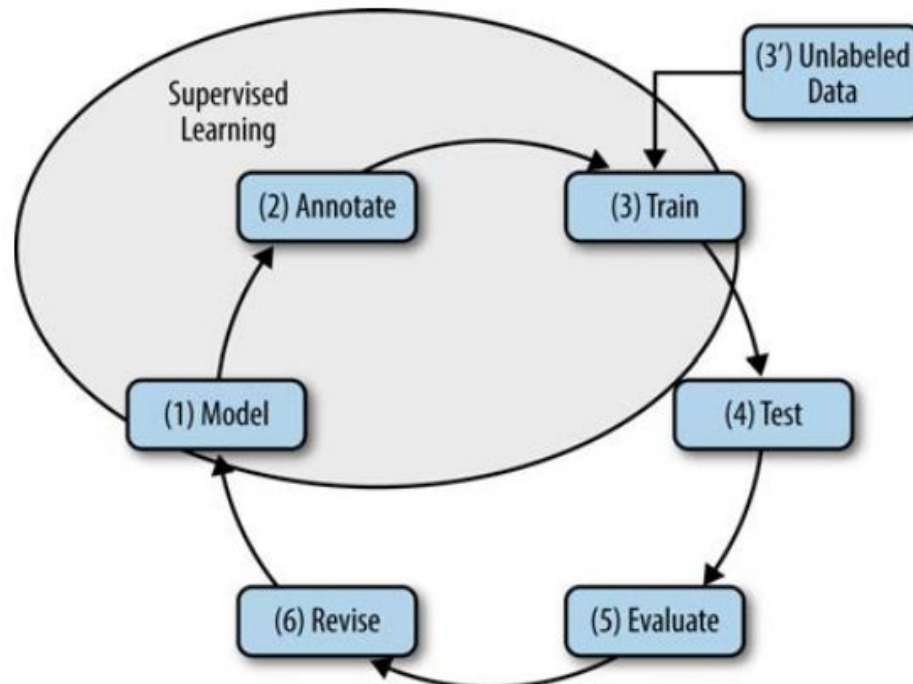


Output in
base alla
forma:



COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO



Il ciclo MATTER

(Pustejovsky and Stubbs (2012) "Natural Language Annotation for Machine Learning". O'Reilly Media.)

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- Il ciclo MATTER:
 - **Model**: descrizione teorica di un fenomeno linguistico
 - **Annotate**: annotazione del corpus con uno schema di annotazione basato sul modello
 - **Train**: addestramento di un algoritmo di ML sul corpus annotato
 - **Test**: test del sistema addestrato su un nuovo campione di dati
 - **Evaluate**: valutazione delle performance del sistema
 - **Revise**: revisione del modello e dello schema di annotazione

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- ANNOTAZIONE
 - aggiunta di informazioni (linguistiche) al testo tramite etichette (*tag*)
 - copre ogni aspetto dell'analisi linguistica
 - rende esplicita e analizzabile dal computer la struttura linguistica implicita nel testo
- SCHEMA DI ANNOTAZIONE
 - repertorio di categorie per l'annotazione: lista di tag e attributi
- LINEE GUIDA DI ANNOTAZIONE
 - documento in cui viene spiegato il *modo* in cui l'annotazione è proiettata sul testo

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO

- Esempio: ***Sentiment Polarity Classification***

subj	Subjectivity: possible values are 0 and 1. A subjective tweet will have subj = 1; an objective tweet subj = 0.
opos	Positive <i>overall</i> polarity: possible values are 0 and 1. A tweet exhibiting positive polarity will have opos = 1; a tweet without positive polarity will have opos = 0.
oneg	Negative <i>overall</i> polarity: possible values are 0 and 1. A tweet exhibiting negative polarity will have neg = 1; a tweet without negative polarity will have neg = 0.
iro	Irony: possible values are 0 and 1. A tweet with an ironic twist will have iro = 1, otherwise iro = 0.
lpos	Positive <i>literal</i> polarity: possible values are 0 and 1. A tweet exhibiting positive <i>literal</i> polarity will have pos = 1; tweet without positive <i>literal</i> polarity will have pos = 0.
lneg	Negative <i>literal</i> polarity: possible values are 0 and 1. A tweet exhibiting negative <i>literal</i> polarity will have neg = 1; tweet without negative <i>literal</i> polarity will have neg = 0.

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO

- Esempio: *Sentiment Polarity Classification*

subj	opos	oneg	iro	lpos	lneg	description and explanatory tweet in Italian
0	0	0	0	0	0	objective <i>l'articolo di Roberto Ciccarelli dal manifesto di oggi</i> http://fb.me/1BQVy5Wak
1	0	0	0	0	0	subjective with neutral polarity and no irony <i>Primo passaggio alla #strabrollo ma secondo me non era un iscritto</i>
1	1	0	0	1	0	subjective with positive polarity and no irony <i>splendida foto di Fabrizio, pluri cliccata nei siti internazionali di Photo Natura</i> http://t.co/GWoZqbxAuS
1	0	1	0	0	1	subjective with negative polarity and no irony <i>Monti, ripensaci: l'inutile Torino-Lione inguaia l'Italia: Tav, appello a Mario Monti da Mercalli, Cicconi, Pont...</i> http://t.co/3CazKS7Y
1	1	1	0	1	1	subjective with both positive and negative polarity (mixed polarity) and no irony <i>Dati negativi da Confindustria che spera nel nuovo governo Monti. Castiglione: "Avanti con le riforme"</i> http://t.co/kIKnbFY7
1	1	0	1	1	0	subjective with positive polarity, and an ironic twist <i>Questo governo Monti dei paschi di Siena sta cominciando a carburare; speriamo bene...</i>

COME SI SVILUPPA UN MODULO TAL

TOKEN	BIO
il	O
capitano	O
della	O
Gerolsteiner	B-ORG
Davide	B-PER
Rebellin	I-PER
ha	O
allungato	O
sulla	O
salita	O

CAT, Content Annotation Tool

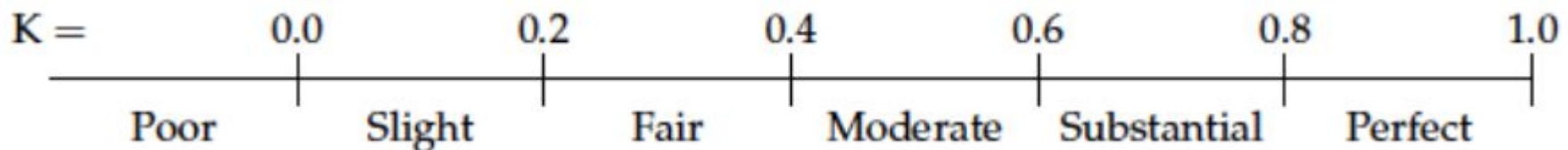
S12	REBELLIN TERZO - Davide Rebellin chiude al terzo posto nella classifica francese Parigi - Bourges e aspetta il gran finale di Coppa del mondo , di cui è l' attuale leader , con la Parigi - Tours di domenica .
S13	A circa 90 chilometri dall' arrivo il capitano della Gerolsteiner ha allungato su uno dei pochi tratti in salita , frazionando il gruppo .
S14	Alla sua ruota si sono portati altri sei corridori che hanno poi disputato lo sprint sul traguardo di Bourges .
S15	La vittoria è andata al francese Jerome Pineau davanti allo svizzero Martin Helmiger .

BRAT

<div> <div>cell type</div> <div>spc cell type</div> <div>anat</div> </div> <p>Characterization of undifferentiated human ES cells and differentiated EBs by antibodies. All monoclonal antibodies were initially selected for their abilities to recognize recombinant proteins in direct ELISAs. A subset were also tested by Western Blot analysis using recombinant proteins and cell lysate to confirm binding to a single epitope. The best clone was later screened for its applications for immunocytochemistry and flow cytometry using various cell lines.</p>	<div> <div>spc</div> <div>anatomy</div> <div>component</div> <div>spc</div> <div>spc</div> <div>gene</div> </div> <p>Human peripheral blood platelets were used for screening mouse anti-human CD9 antibody.</p>
<div> <div>c line</div> <div>spc</div> <div>spc</div> <div>gene</div> <div>gene</div> <div>gene or protein</div> </div> <p>MCF-7 cells were used for screening mouse anti-human E-Cadherin and PODXL (podocalyxin-like) antibodies.</p>	<div> <div>c line</div> <div>spc</div> <div>spc</div> <div>gene</div> <div>gene or protein</div> </div> <p>MG-63 cells were used for screening mouse anti-human GATA1 (GATA binding protein 1) antibody.</p>

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- **Inter-Annotator Agreement (IAA)** = accordo tra almeno 2 annotatori sullo stesso testo
 - consistenza dell'annotazione
 - plausibilità cognitiva del modello
 - un ampio accordo tra gli annotatori è considerato garanzia della validità di tale schema e dei dati annotati
 - K di Cohen (annotatori = 2) o di Fleiss (annotatori > 2)



COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- Dati necessari:
 - di training (*training set*): dati annotati per l'addestramento del modello
 - di test (*test set*): dati NON annotati, diversi da quelli di training, su cui applicare il modello addestrato
 - di valutazione (*gold standard*): dati del test annotati su cui valutare le performance del modello addestrato

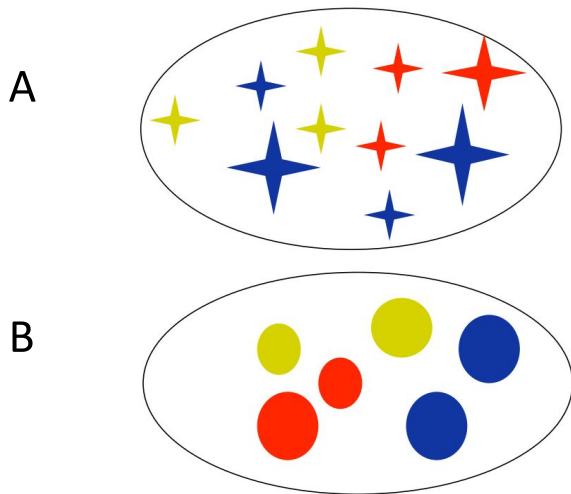
MACHINE LEARNING: TRAINING E TEST

- TRAINING:
 1. Selezionare una serie di documenti da usare come training
 2. Annotare ciascun token nei documenti
 3. Identificare l'approccio migliore
 4. Creare un classificatore che preveda le etichette
- TEST:
 1. Usare una serie di documenti non usati nel training
 2. Lanciare il classificatore per etichettare ciascun token
 3. Avere in output i documenti annotati
 4. Valutare sugli stessi documenti manualmente annotati

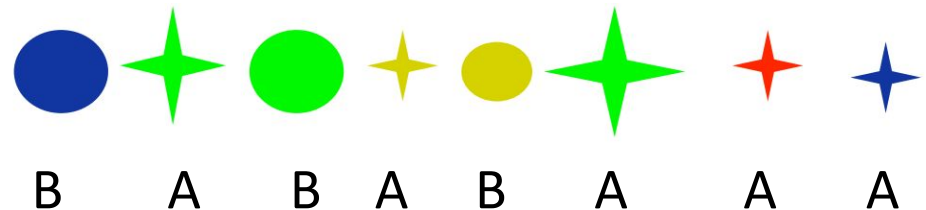
COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO, esempio
 - CLASSIFICAZIONE: dato un insieme di classi predefinite determinare a quale classe appartiene una certa entità

Input (training):

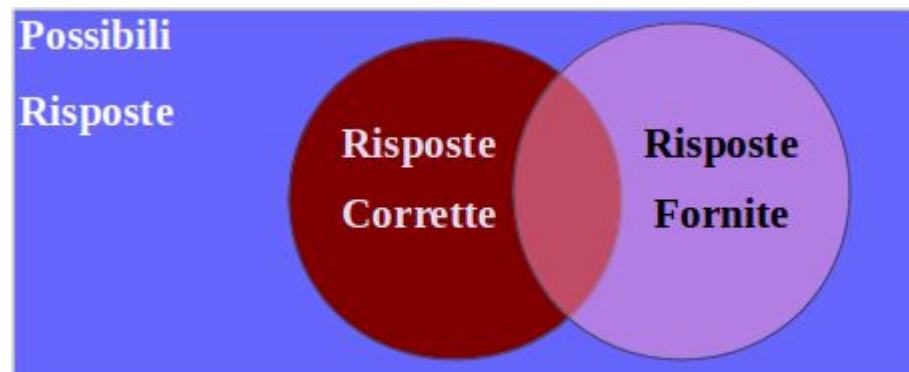


Classificazione di nuovi dati (test):



COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- **VALUTAZIONE:** analisi quantitativa delle prestazioni del modello
 - confronto dell'output del modello sui dati di test con il gold standard



COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- **VALUTAZIONE:** analisi quantitativa delle prestazioni del modello
 - uso di metriche standard: ACCURACY

$$\text{ACCURACY} = \frac{\text{\#risposte corrette}}{\text{\#risposte fornite}}$$

Esempio:

- 150 frasi annotate nel test
- 120 frasi annotate con sentiment corretto
- accuracy = $120/150 = 0,8$ (80%)

DEMO ONLINE

Pipeline di Stanford CoreNLP:

- sei lingue: inglese, francese, tedesco, spagnolo, cinese, arabo
- moduli: sentence splitting, lemmatizzazione, PoS tagging, parsing, NER, coreference, sentiment, relazioni
- demo online: <http://corenlp.run/>

DEMO ONLINE

HeidelTime:

- riconoscimento e normalizzazione espressioni temporali
- 13 lingue tra cui l'italiano
- esportazione in XML
- demo online: <https://heidelttime.ifi.uni-heidelberg.de/heidelttime/>



GRAZIE!

Email: rachele.sprugnoli@unicatt.it

Twitter: [@RSprugnoli](https://twitter.com/RSprugnoli)



COME SI SVILUPPA UN MODULO TAL


- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- **VALUTAZIONE:** matrice di confusione

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- **VALUTAZIONE**: analisi quantitativa delle prestazioni del modello
 - uso di metriche standard: **PRECISION**, misura il rapporto tra le entità correttamente riconosciute dal sistema ed il totale delle entità riconosciute

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

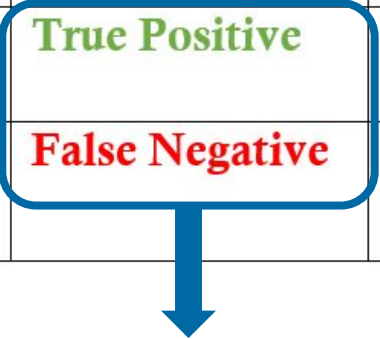


$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- **VALUTAZIONE**: analisi quantitativa delle prestazioni del modello
 - uso di metriche standard: **RECALL**, misura il rapporto tra le entità correttamente riconosciute dal sistema ed il totale delle entità corrette

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative



$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- **VALUTAZIONE**: analisi quantitativa delle prestazioni del modello
 - uso di metriche standard: **F-MEASURE**, media armonica tra precision e recall

$$\text{F-MEASURE} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

COME SI SVILUPPA UN MODULO TAL

- Sistemi di apprendimento automatico – MACHINE LEARNING (ML)
- ML SUPERVISIONATO
- **VALUTAZIONE:** analisi quantitativa delle prestazioni del modello
 - Esempio:

		ACTUAL (gold standard)	
		Positive	Negative
PREDICTED (test set)	Positive	70 (TP)	15 (FP)
	Negative	30 (FN)	45 (TN)

- Precision: $70 / (70+15) = 70 / 85 = 0,82$

- Recall: $70 / (70+30) = 70 / 100 = 0,70$

- F-measure: $2 * 0,82 * 0,7 / (0,82 + 0,70) = 0,75$