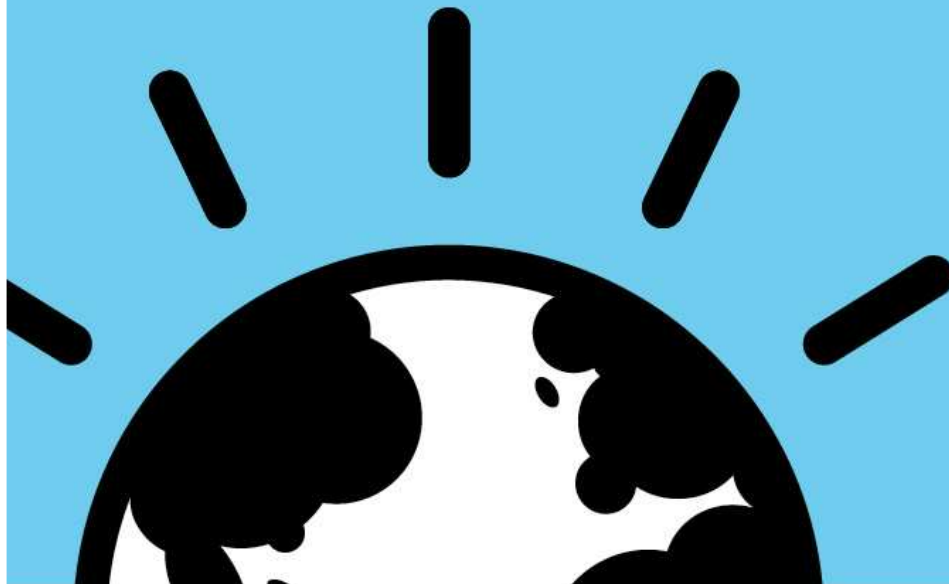


Siete pronti per un pianeta più intelligente?

Guido Vetere

Chiedilo a Watson!

IBM DeepQA e le sue applicazioni nella Pubblica Amministrazione



FORUM PA 2011

Roma, 9 - 12 maggio



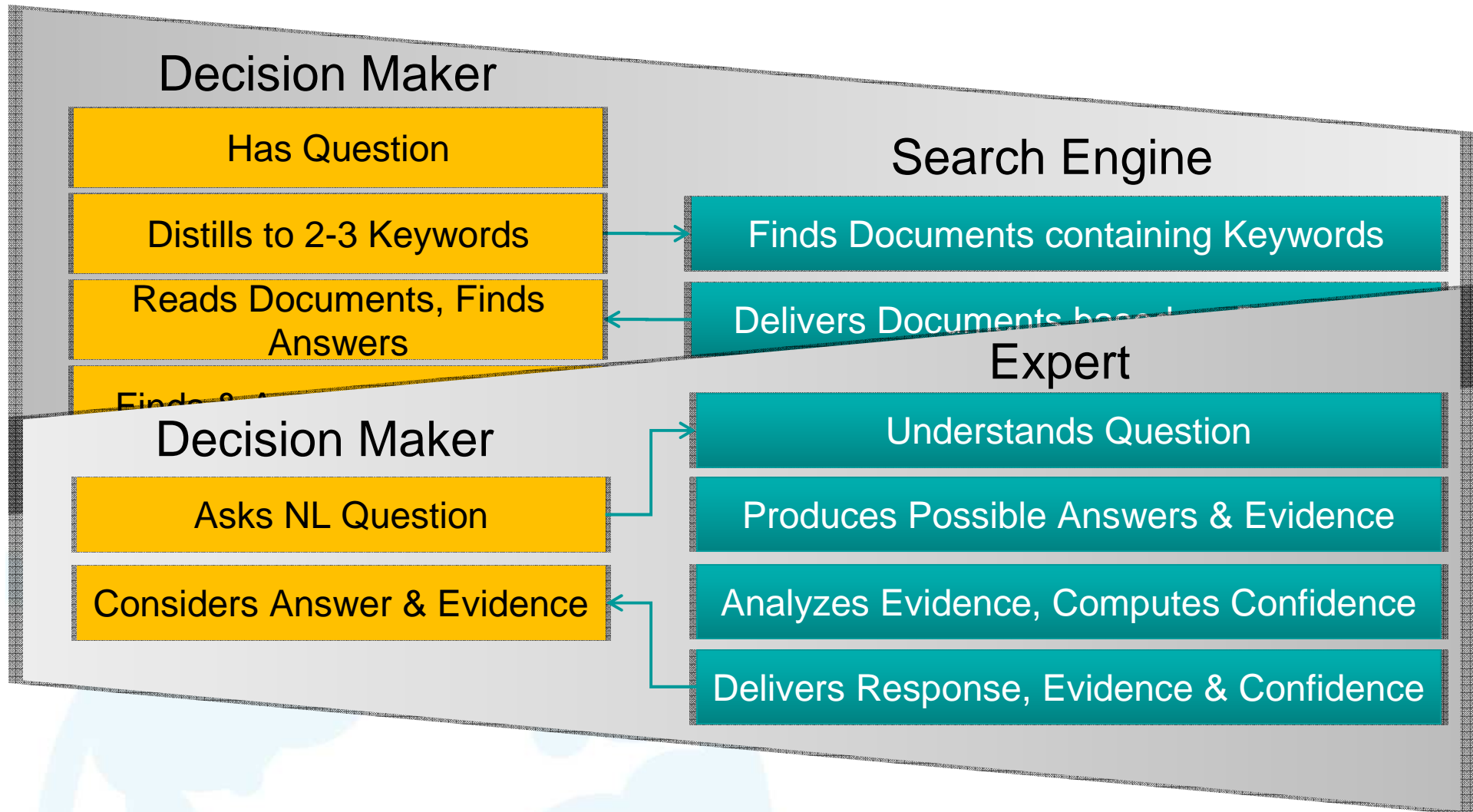
Chi sono

- Direttore del Centro Studi Avanzati IBM di Roma
- Mi occupo di ricerca
 - Rappresentazione della conoscenza e ontologia
 - Ragionamento automatico e integrazione di informazione
 - Linguaggio naturale

La presentazione

- La sfida del Question Answering
- Il sistema IBM DeepQA (*Watson*)
- Question Answering e Pubblica Amministrazione

Cos'è il Question Answering e perché è interessante



Siete pronti per un pianeta più intelligente?



QA in azione: Watson a *Jeopardy!*



([clip](#))

Ma era poi così difficile?

- Where was X born?

One day, from among his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of Einstein's birthplace.

Person	Birth Place
A. Einstein	ULM

- X ran this?

If leadership is an art then surely Jack Welch has proved himself a master painter during his tenure at GE.

Person	Organization
J. Welch	GE

Per i computer ci sono cose facili ...

$$\ln((12,546,798 * \pi)) ^ 2 / 34,567.46 = \mathbf{0.00885}$$

Select *Payment* where *Owner*="David Jones" and *Type(Product)*="Laptop",

Owner	Serial Number
David Jones	45322190-AK

Invoice #	Vendor	Payment
INV10895	MyBuy	\$104.56

Serial Number	Type	Invoice #
45322190-AK	LapTop	INV10895

David Jones

 David Jones

=

Dave Jones

 David Jones

≠

... e meno facili

“Stamattina ho sparato ad un elefante in pigiama. Mi chiedo come abbia fatto ad entrarci ..”



Siete pronti per un pianeta più intelligente?



	Human	Watson (programmed by humans)
Question Understanding	Seemingly Effortless. Almost instantly knows what is being asked, what is important and how it applies – very quickly resolves focus, relevant parts, references, hints, puns, implications, etc.	Hugely Challenging. Has to be programmed to analyze enormous numbers of possibilities to get just a hint of the relevant meaning. Very difficult due to variability, implicit context, ambiguity of structure and meaning in language.
General Language Understanding	Seemingly Effortless. Powerful, general, deep and fast in understanding language – reading, experiencing, summarizing, storing knowledge in natural language. This information is written for human consumption so reading and understanding what it says is natural for humans.	Hugely Challenging. Answers need to be determined and justified in natural language sources like news articles, reference texts, plays, novels etc. Watson must be carefully programmed and intensely trained to deeply analyze even just tiny subsets of language effectively. Very different from web search that returns documents containing the question words ranked by popularity. Rather, must find a precise answer and understand enough of what it read to know if and why a possible answer may be correct.
Self-Knowledge (Confidence)	Seemingly Effortless. Most often, and almost instantly, humans know if they know the answer.	Hugely Challenging. 100's of algorithms are used to find and analyze 1000's of written texts for many different types of evidence, then the results are combined, scored and weighed for their relative importance – how much they justify a candidate answer.
Breadth of Knowledge	Limited by self-contained memory. Estimates of >1000's of terabytes are all much higher than Watson's memory capacity. Ability to flexibly understand and summarize human relevance (i.e., compress) means that humans' raw input capacity is even higher. But what any person decides to do with their memory varies of course.	Limited by self-contained memory. Roughly about 0.5 to 1 million books worth of content memory. Weaker ability to meaningfully understand and summarize human-relevant components. Does not, of course, include a full life experience.
Processing Speed	Due to relatively instant language abilities, highly associative, highly flexible memory and speedy recall, generally much faster to grasp question, determine if it knows and to get the answer.	Hugely Challenging. On 1 CPU Watson can take over 2 hours to determine if it confidently knows the answer to a typical Jeopardy! question. Watson must be parallelized, perhaps in ways similar to the brain, to use 1000's of compute cores to compete against humans in the 3–5 second range.
Reaction Speed	Slower raw reaction speed but determines confidence and answer faster and with less effort. Has the ability to listen to clue and anticipate when to ring in, providing humans with the fastest absolute possible response time.	More consistently can deliver a fast reaction time but ONLY IF and WHEN can determine high enough confidence in time to ring-in. Not able to anticipate when to ring-in based on listening to clue, which gives fastest possible response time to humans.
Compute Power	Requires 1 brain that fits in a shoe box, can run on a tuna-fish sandwich and be cooled with a hand-held paper fan.	Hugely Challenging: Needs >2500 compute cores requiring 80Kw of power and 20 tons of cooling (8–10 refrigerators worth in size and space)
Betting	Slower, less precise.	Faster more accurate calculations.
Emotions	Yes. Can slow down and /or confuse processing.	No. Does NOT get nervous, tired, upset or psyched out. (but the Team does !!)

Siete pronti per un pianeta più intelligente?



La Grande Sfida



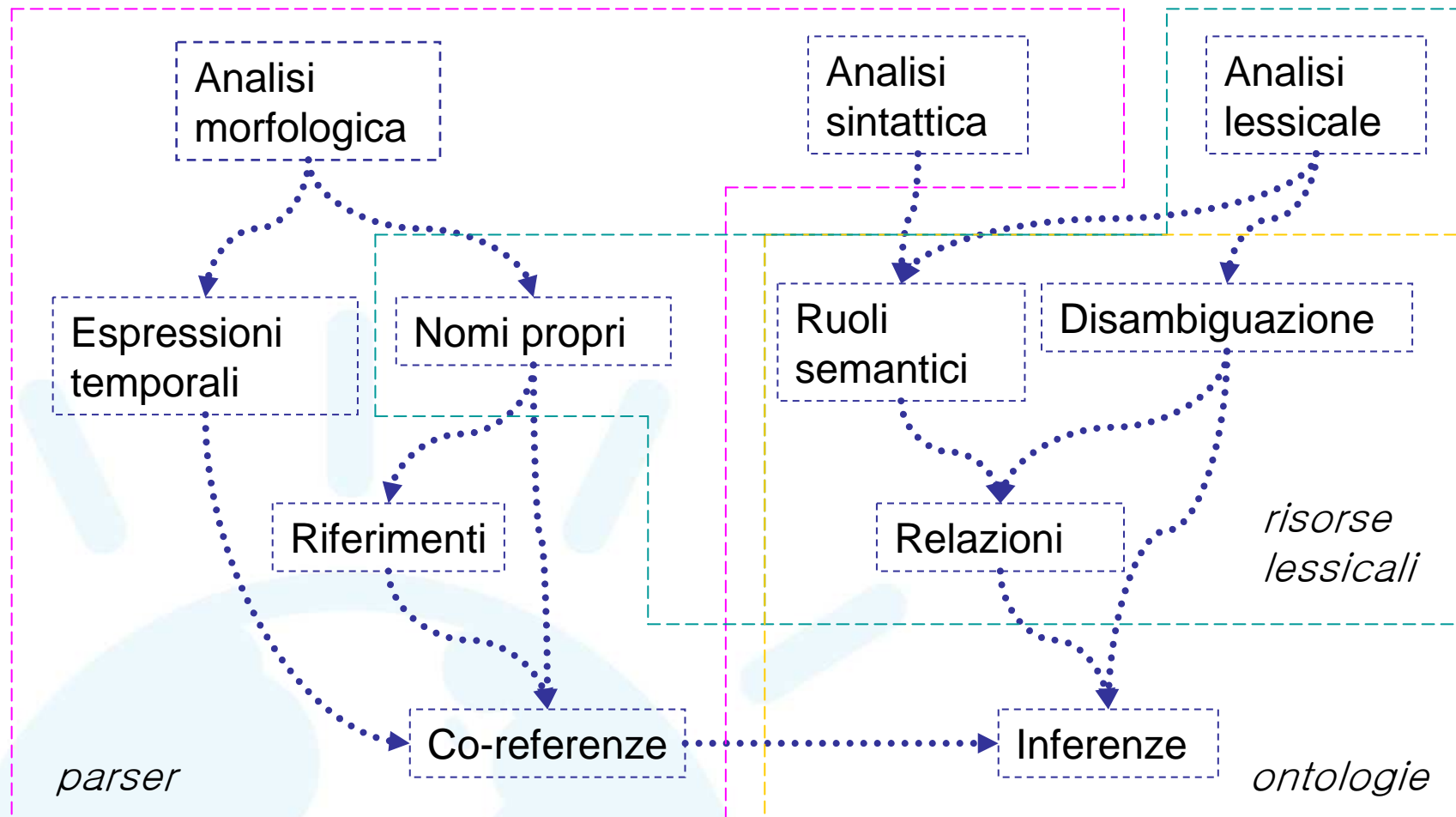
([clip](#))

Il linguaggio naturale

- Il linguaggio naturale è la principale forma di codifica e di accesso della conoscenza
 - Dati non strutturati > 80% del totale disponibile
- Qualsiasi progresso nelle tecnologie del linguaggio può avere grandi ricadute pratiche
 - Semantic Search
 - Knowledge Management
 - Business Analytics
 - Question Answering



Una mappa delle tecnologie della lingua



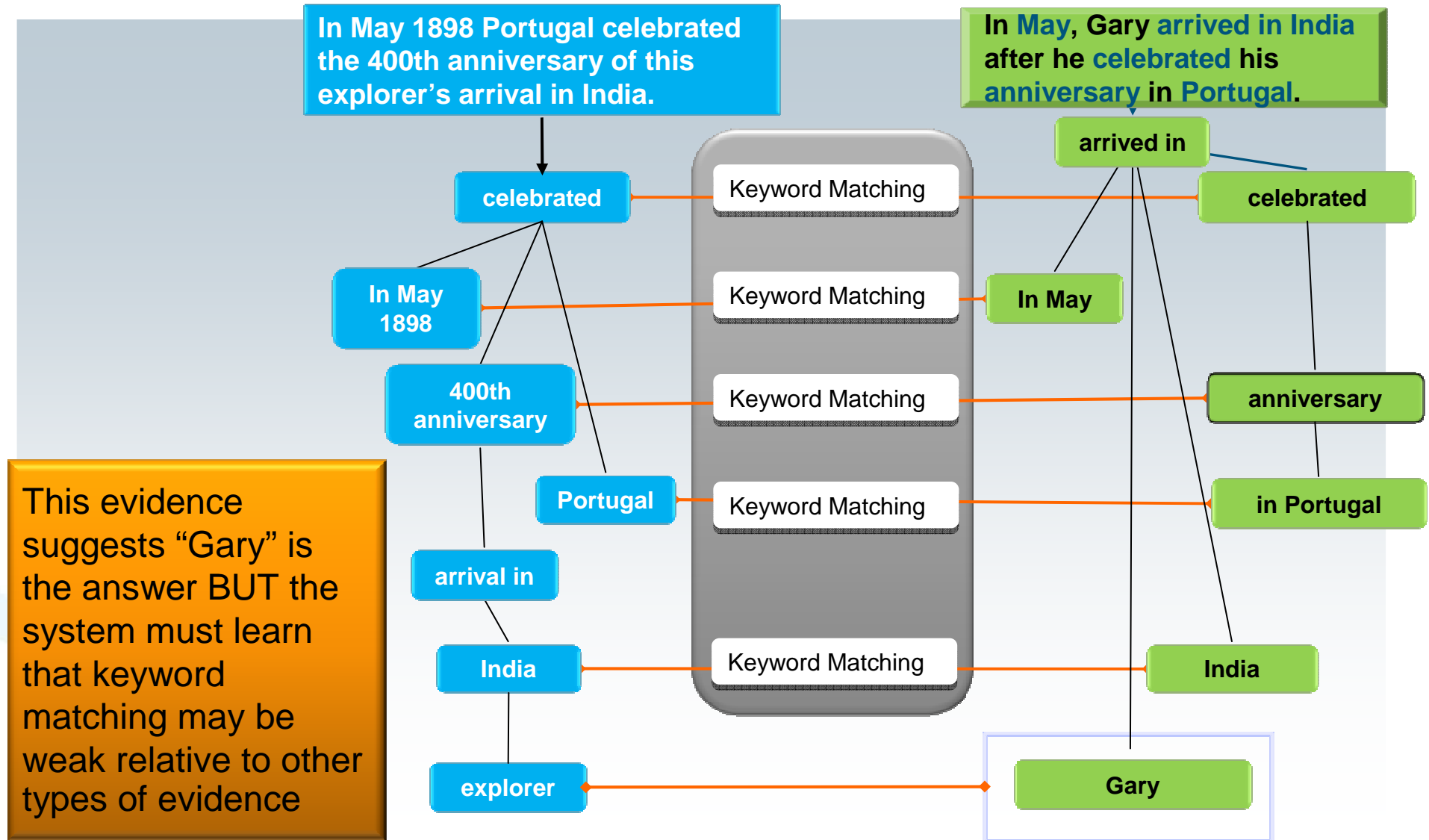
Siete pronti per un pianeta più intelligente?



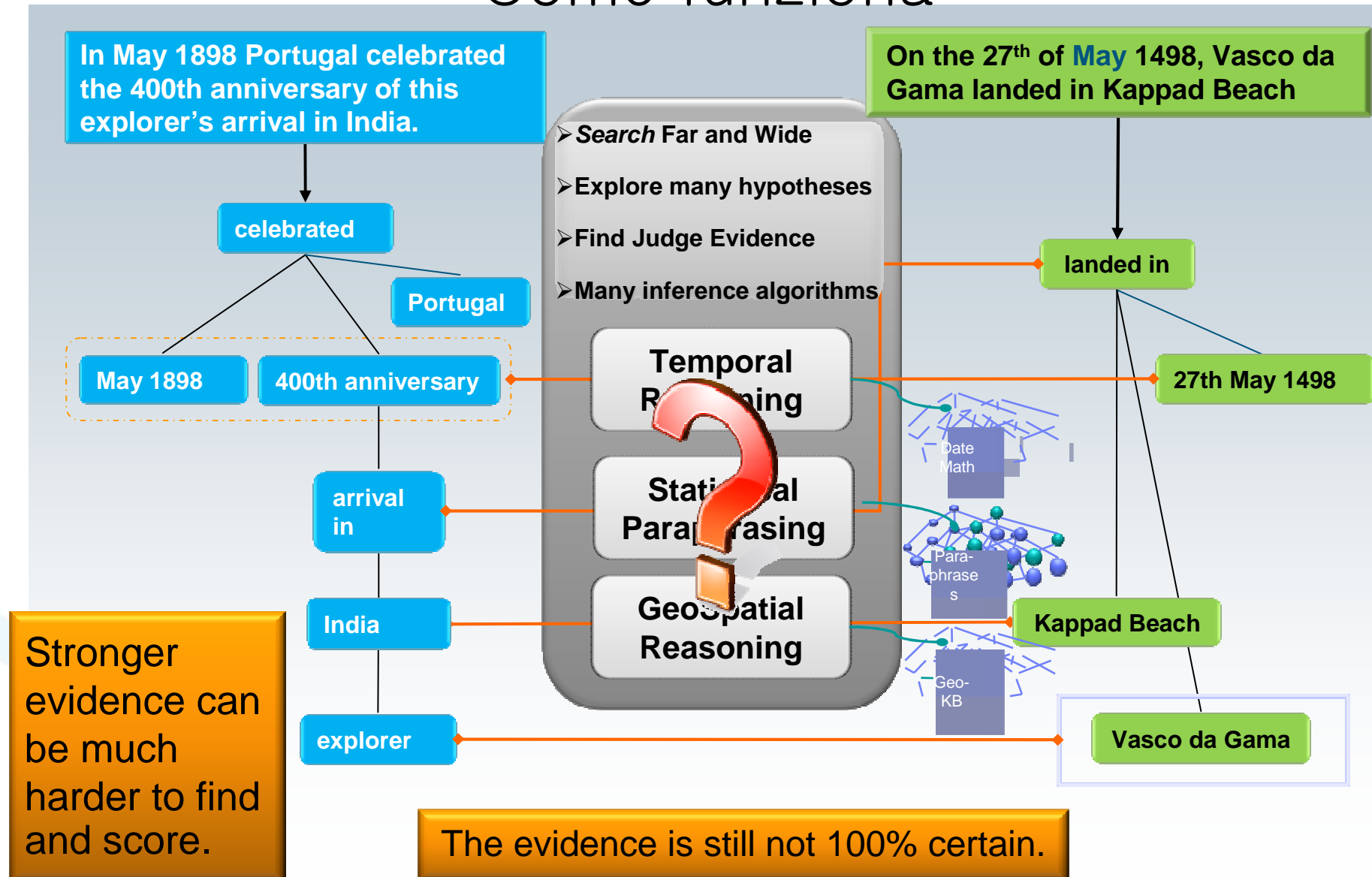
Dentro Watson



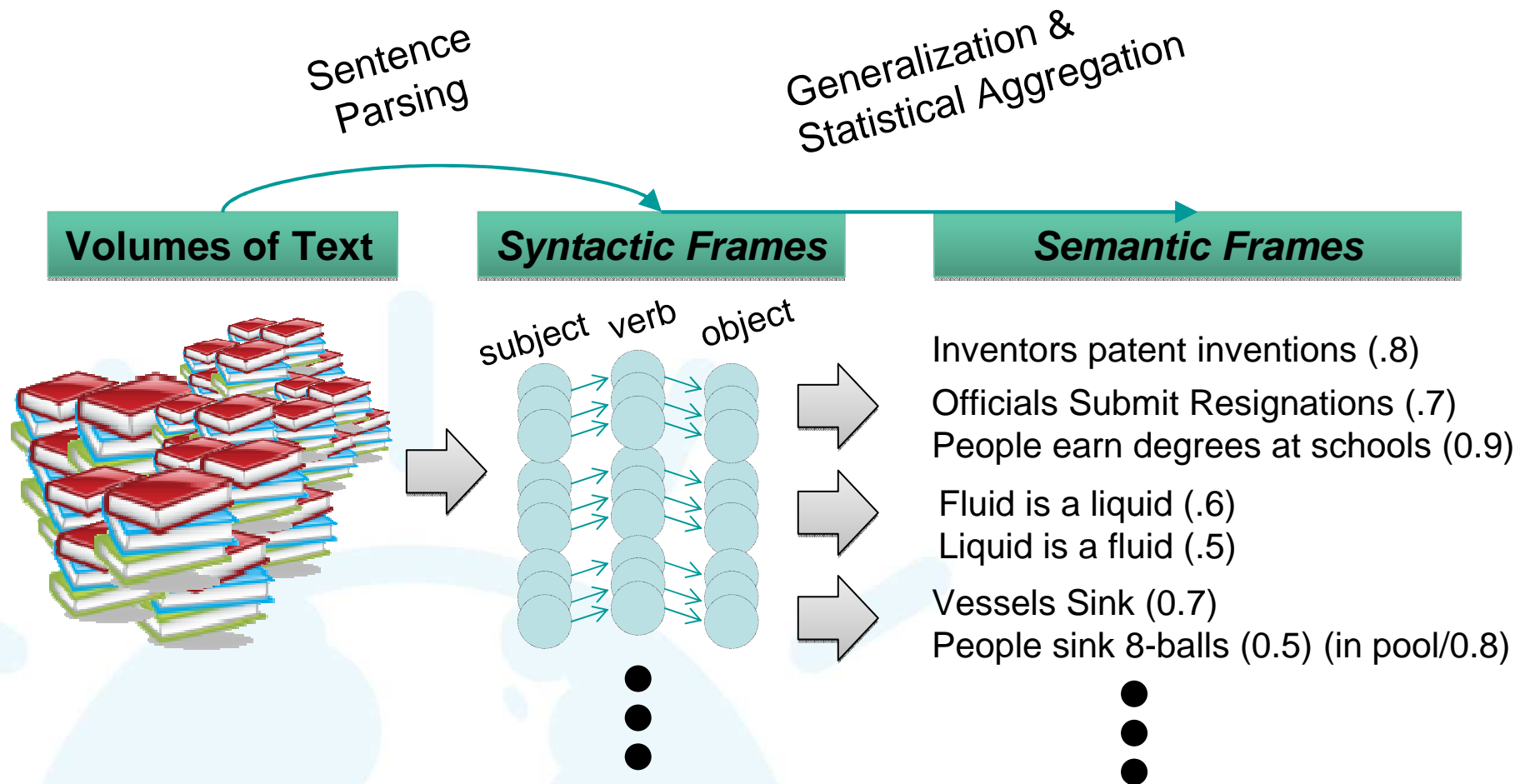
Come non funziona



Come funziona

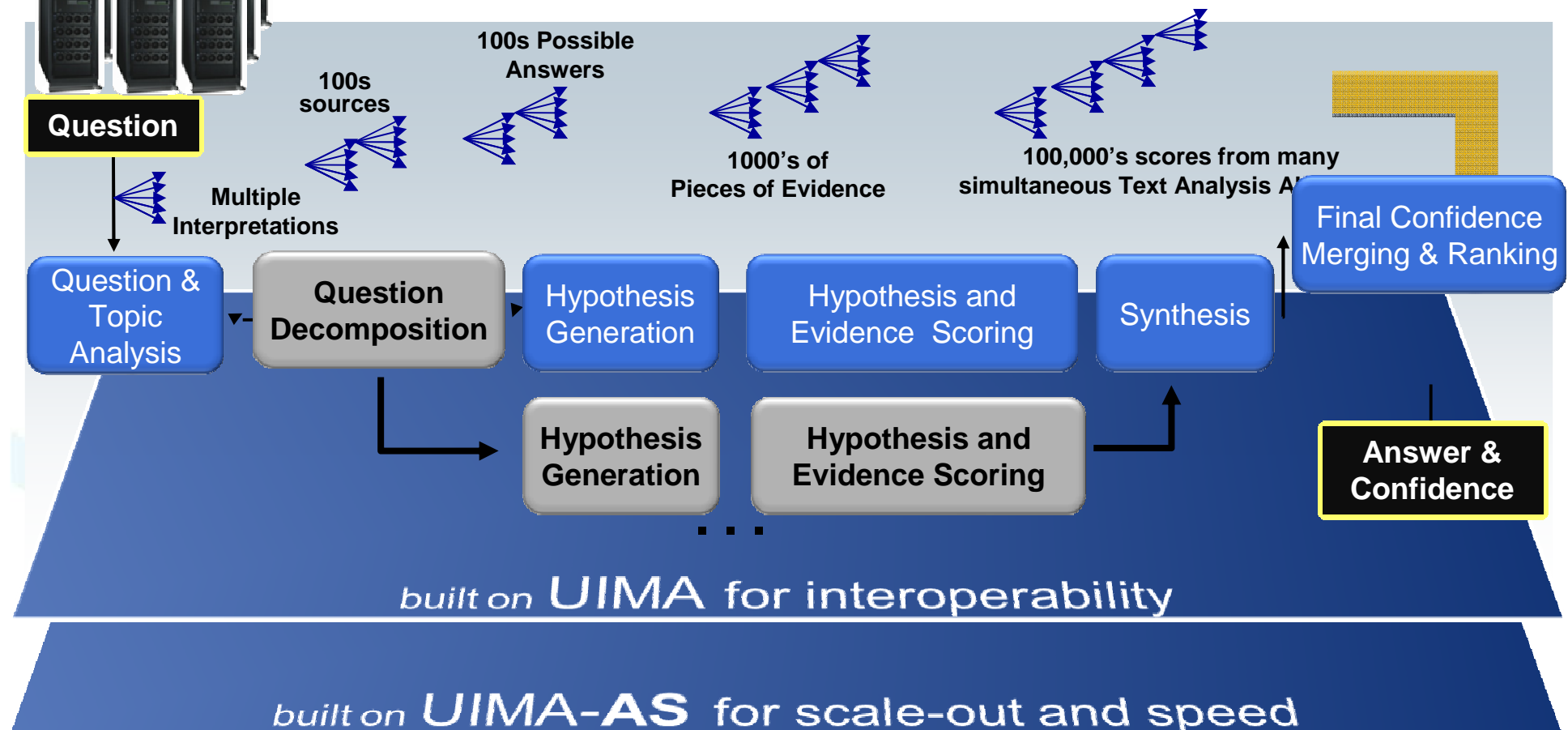


Il processo di apprendimento



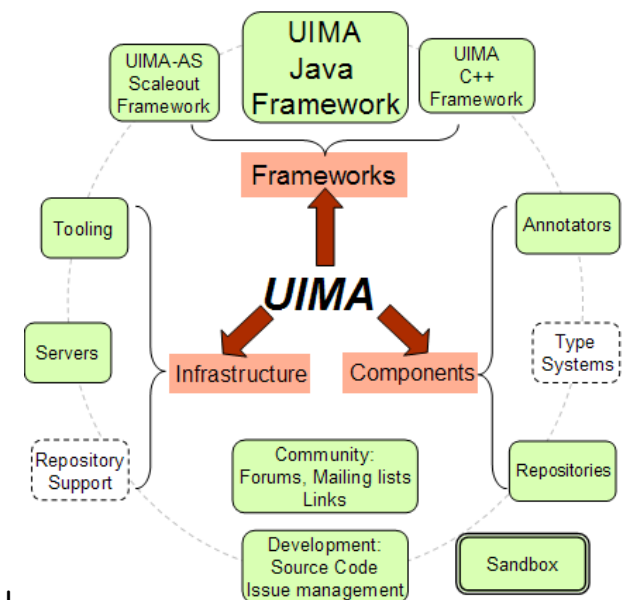
La tecnologia

*Generates and scores many hypotheses using a combination of 1000's **Natural Language Processing, Information Retrieval, Machine Learning and Reasoning Algorithms.***



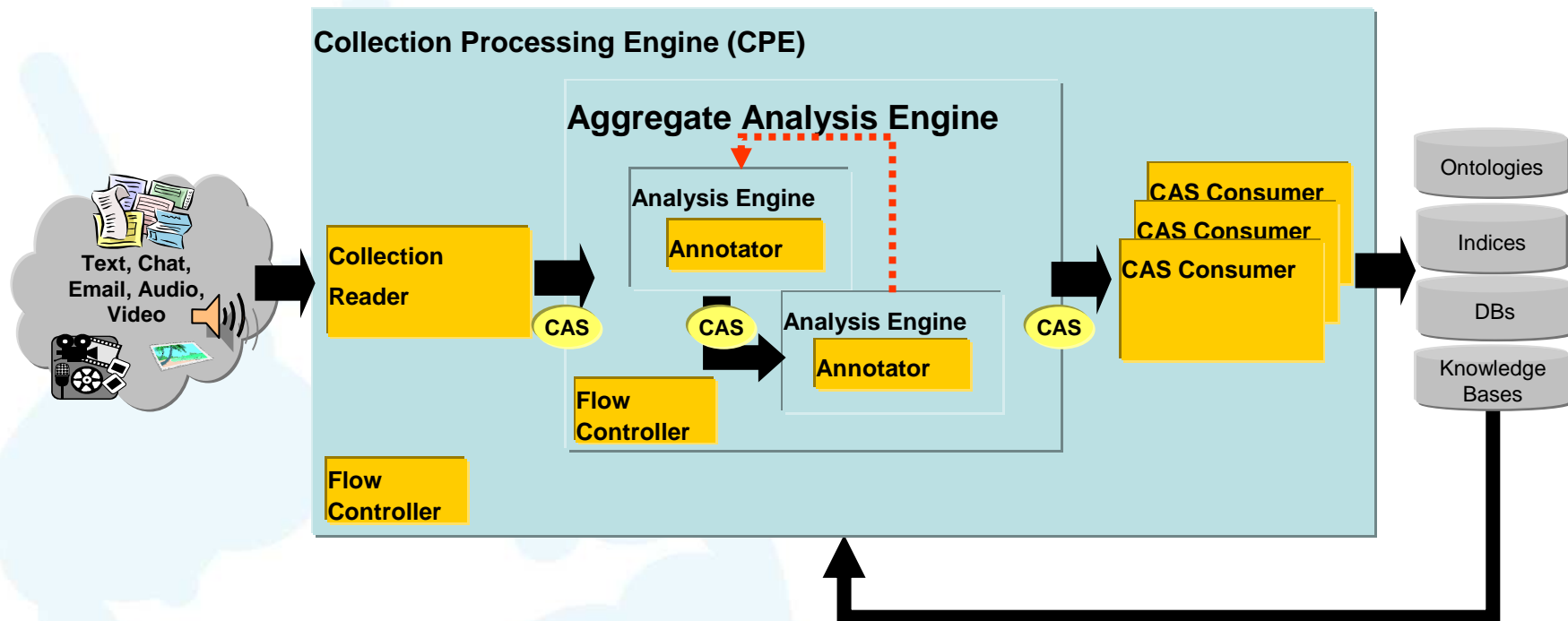
UIMA, il cuore aperto di Watson

- Apache Unstructured Information Management Architecture (uima.apache.org)
- Framework (Java, C++) che consente di sviluppare e integrare componenti analitiche per informazione testuale
- Esempi:
 - Identificazione di entità (persone, luoghi, organizzazioni, ...)
 - Identificazione di relazioni (lavora-per, situato-in, ...)
- Implementazioni
 - Apache Foundation (OS)
 - IBM Content Analytics with Enterprise Search (include Omnifind)
 - Annuncio: Aprile 2011



UIMA Common Analysis Structure

- L'architettura UIMA su basa su componenti standardizzate (Common Analysis Structure, CAS).
- Le CAS incorporano gli algoritmi analitici e forniscono loro un set di interfacce standardizzate
- Le CAS si possono comporre tra loro per implementare euristiche sofisticate



Siete pronti per un pianeta più intelligente?



Watson al lavoro

Enterprise Content Management

- Advanced Case Management
- Content Analytics
- Document Imaging and Capture
- Information Lifecycle Governance
- Social Content Management



Business Analytics

- Analytic Applications
- Business Intelligence
- Predictive Analytics
- Financial Performance Management
- Governance, Risk & Compliance
- Web Analytics



Information Integration & Federation

- Information Integration
- Master Data Management
- Data Warehousing
- Big Data and Streams



Data Management

- Database Software
- Database Management Tools



Information Governance

- Data Lifecycle Management
- Data Security and Privacy



Potenziali applicazioni



Healthcare / Life Sciences: Diagnostic Assistance,
Evidenced-Based, Collaborative Medicine

Tech Support: Help-desk, Contact Centers



Enterprise Knowledge Management and Business Intelligence

Government: Improved Information Sharing
and Security



Grazie

Guido Vetere

Manager & Research Coordinator, IBM Center
for Advanced Studies Rome

IBM Italia S.p.A.
via Sciangai 53, 00144 Rome,
Italy

mail: gvetere@it.ibm.com
phone: +39 06 59662137
mobile: +39 335 7454658