José Manuel Gómez Pérez

RETELE – Sept. 19th, 2017

EXPERT SYSTEM
SEMANTIC INTELLIGENCE

ABOUT US | COGITO | SOLUTIONS | PRODUCTS | CUSTOMERS | PARTNERS | NEWS | INVESTOR RELATIONS

About us

Home » About us

COGITO

Listed on AIM Exchange in February 2014

Expert System develops software that understands the meaning of written language.

When we talk about the early days of Expert System, we often lead with what has become an inside joke: "we were a start-up but didn't know it." As three university colleagues with a dream, we were determined to demonstrate that great software could be created in Italy.

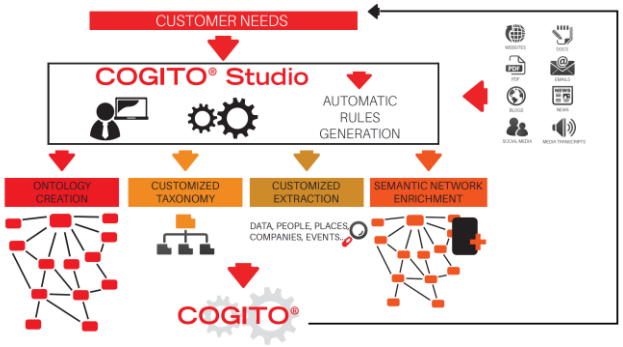
At the time, the convergence of linguistics and technology was something only being talked about in research institutions or academia. There were no start-up incubators or angel investors; we were young and bootstrapping the business with what we had, doing ordinary tech work during the day, and pursuing our vision at night.

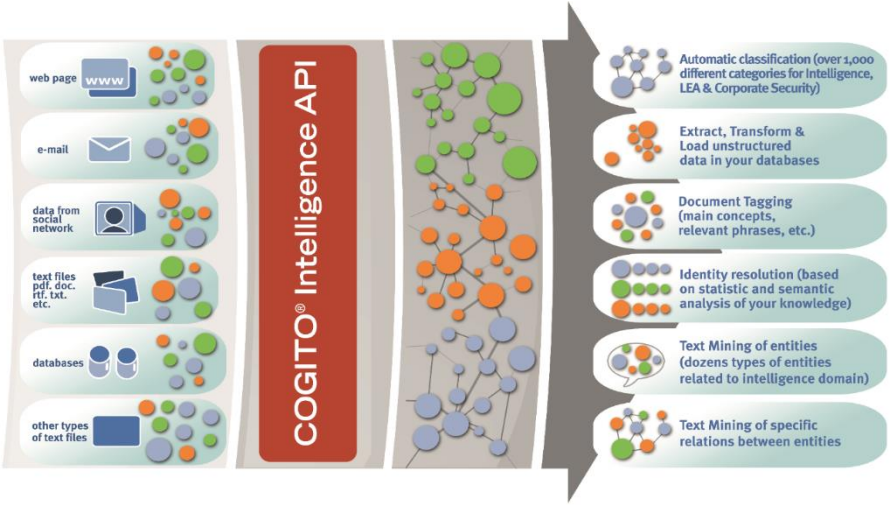
After licensing our early technology to Microsoft, we were able to fully extend our vision to developing software that could understand the meaning and context of language. The effort produced one of the first semantic analysis platforms and led to our patented **Cogito technology**.

BANKING & INSURANCE
PUBLIC ADMINISTRATION
TELCO
MEDIA & PUBLISHING
ENERGY, OIL & GAS

Expert System's COGITO

14 languages natively supported





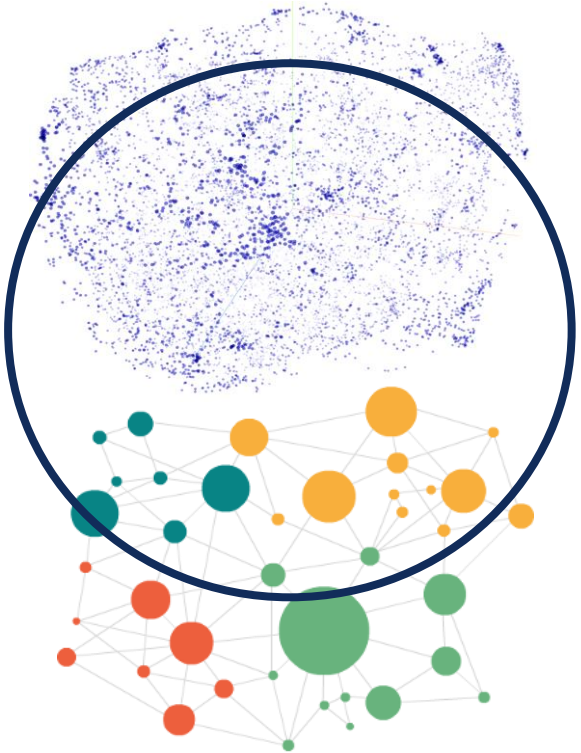
- Based on **Sensigrafo, a monolingual knowledge graph** containing word definitions, related concepts and linguistic information
- Main entities include **syncons** (concepts), **lemmas** (canonical representation of a word) and **relations** (properties, taxonomical, polysemy, synonymy…)
  - 301,582 syncons
  - 401,028 lemmas
  - 80+ relation types that yield ~2.8 million links
- **Internal representation** leverages external resources, both general and sector-specific, e.g. Wikidata, RAE….
- **Word-sense disambiguation**, based on the context of a word in Sensigrafo
- **Categorization** and **extraction** supported through **Sensigrafo plus lexical-syntactic rules**

Challenges and opportunities: Multilingualism and reuse

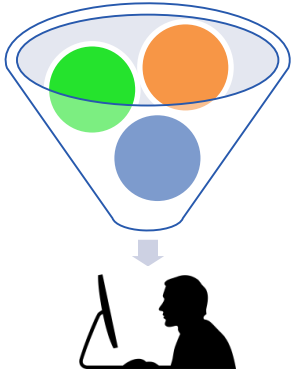
- Expert System's fast internationalization has raised the **challenge of creating from scratch new monolingual resources** (Sensigrafos and rules) for the new languages
- **… or rather, achieve multilingualism in a cost-effective manner**, while maintaining high accuracy and reducing time to market
- **Systematic MT is not the solution,** due to domain, business and territorial nuances
- However, many of the projects in the new languages, e.g. Spanish back in the day, are **conceptually similar to past projects in well-supported languages**
- Our goal is to **enable the reuse of in-house** semantic and linguistic resources, **quickly evolving incipient Sensigrafos**

Vecsigrafo – A hybrid statistic and symbolic knowledge representation

- Knowledge embedded in document corpora
- Statistic induction
- Broad, flexible, scalable
- Good for POS tagging, dependency parsing, semantic relatedness
- Lack of true understanding of real-world semantics and pragmatics

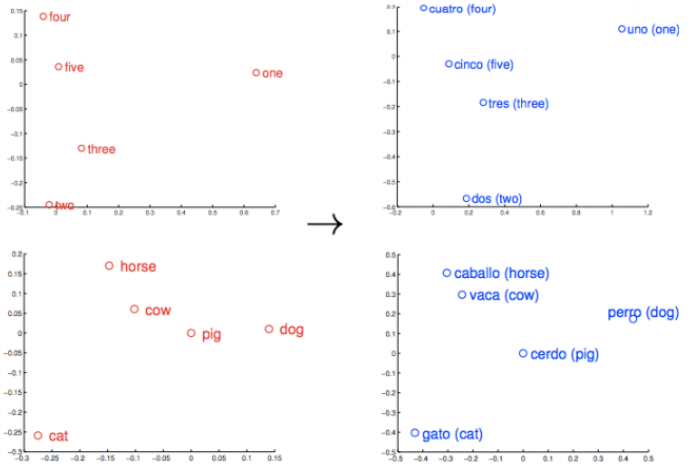


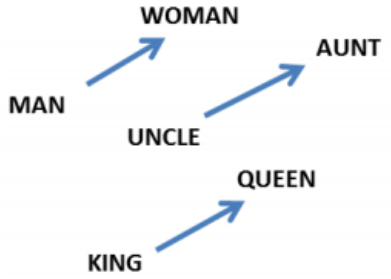
- Knowledge encoded in the mind of the expert
- Structured knowledge base
- Deep, but rigid and brittle
- Good for logical deduction and explanation
- Human is a bottleneck: hand-engineered features and powerful modeling tools needed

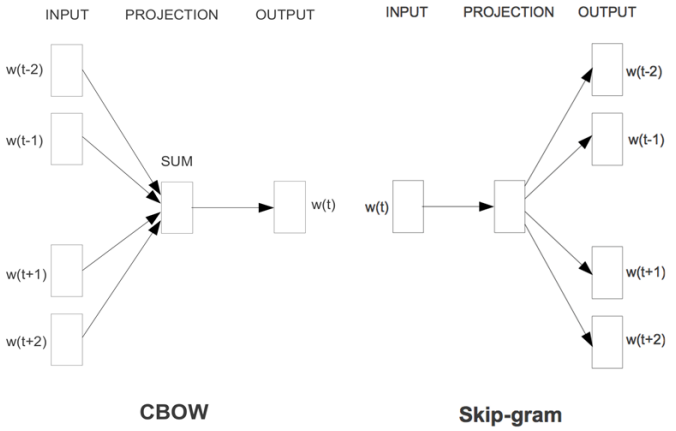


Distributed Word representation - Word2vec

- Word2vec represents words in a vector space, making natural language computer-readable
- Neural word embeddings enable word similarity and relatedness based on vector arithmetic, e.g. cosine similarity
- **Semantic portability across languages**







Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Preliminary experiments

- **Two parallel corpora**, focused on English and Spanish: **Europarl** and **UN**
- **Tokenized, lemmatized and disambiguated** with COGITO
- Learned monolingual models and a transformation between their vector spaces
- **Deeplearning4j** with Skip-gram, minFreq 10, vector dimensionality 400
- Switched to **TensorFlow** and **Swivel** for better vectorization time (~16x and ~20x speedup for 15 epochs)
- Three main use cases
  - **Interlinking** monolingual sensigrafos across different languages
  - Automatic identification of **disambiguation errors** in COGITO
  - **Curation and identification of modeling gaps** in Sensigrafo

Corpus	Sentences	Spanish words	English words
Euparl	1,965,734	51,575,748	49,093,806
UN	21,911,121	678,778,068	590,672,799

Conclusions and needs

- Very promising results
  - See **(Denaux and Gomez-Perez, 2017)** for a deeper look into our work
- But we need **more**:
  - High quality knowledge bases on **vertical domains of economic and scientific interest**
    - **Who is accountable for curating this?**
  - Available large corpora (~10^8 sentences, ideally), in multiple languages
  - **Not so much about general knowledge (commodity)** but corpora rich in specific, vertical domains
  - Work on **incremental methods** for the generation of vector embeddings to speed-up training and increase timeliness

BANKING & INSURANCE
PUBLIC ADMINISTRATION
TELCO
MEDIA & PUBLISHING
ENERGY, OIL & GAS

Jose Manuel Gomez-Perez, PhD
Director R&D
jmgomez@expertsystem.com

Denaux R, Gomez-Perez JM. ***Towards a Vecsigrafo: Portable Semantics in Knowledge-based Text Analytics.*** *To appear in proceedings of the Intl. Workshop on Hybrid Statistical Semantic Understanding and Emerging Semantics (HSSUES), collocated with the 16th Intl. Semantic Web Conference (ISWC), Vienna, 2017.*

linkedin.com/company/expert-system

twitter.com/Expert_System

info@expertsystem.com