

ATOKA.IO

EL GRAFO MÁS GRANDE DE INFORMACIÓN EMPRESARIAL OFICIAL Y NO-OFICIAL EN ITALIA



Javier Paniagua es egresado del DEI, UC (Ing. Informática) y PhD en ICT del DISI, Universidad de Trento.

Trabaja en R&I en SpazioDati, una joven empresa italiana especializada en integración, análisis e interpretación de datos empresariales. Ha participado en varios proyectos de los programas FP7 y H2020 de la Unión Europea en las áreas de multimedia, GIS, NLP y Knowledge Representation. Actualmente es Lead Engineer en el equipo encargado de mantener Atoka.io, el directorio de información empresarial líder en Italia, y Project Manager en el Consorcio euBusinessGraph, encargado de extender Atoka.io para operar en nuevos mercados.

Jueves 31 de Mayo – 15.30hs – Sala Manfred Stark

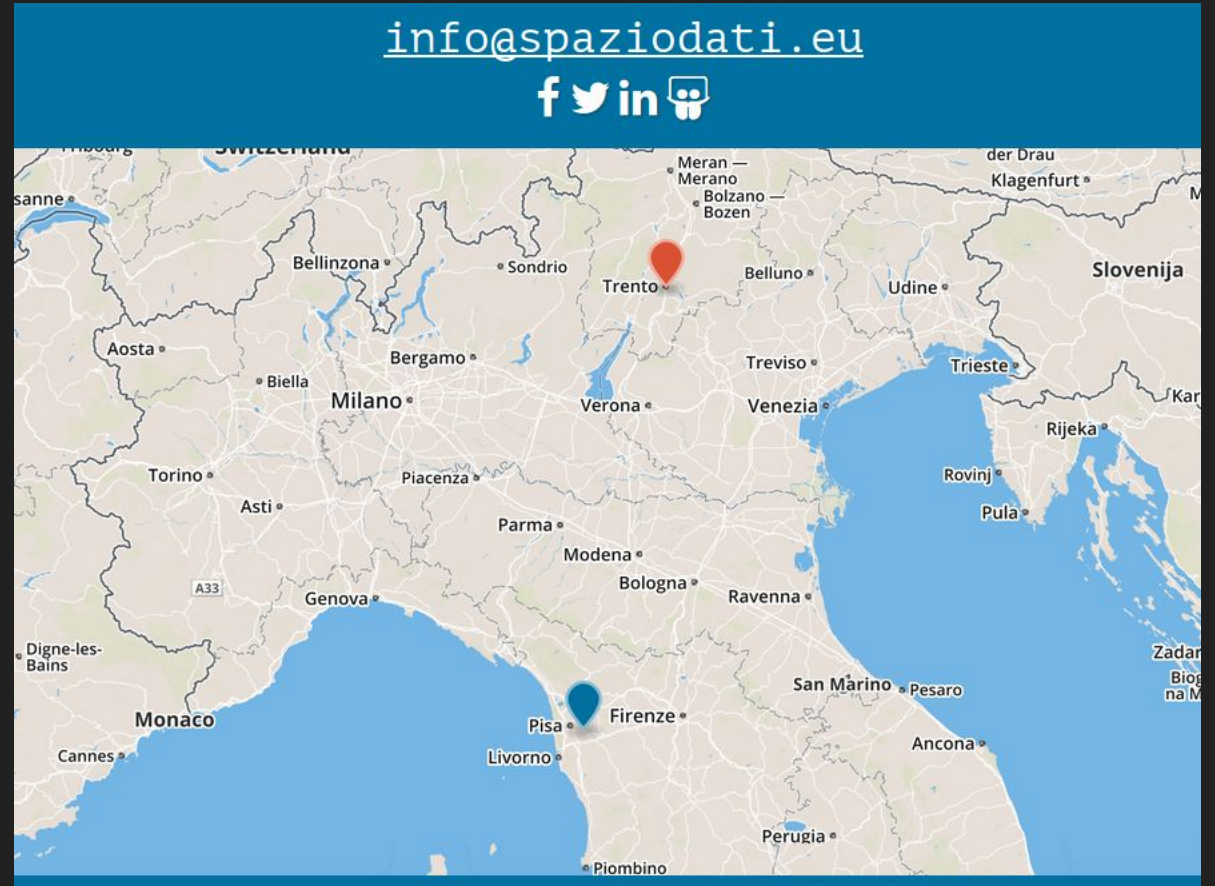


Bio

- Egresado de la UC Asunción en el 2005 (Ing. Informática)
- Algunos años de experiencia freelance
- Sistemas de tasación para telefonía celular
- Pasantía + PhD en UniTN en 2015 (ICT: personal photography)
- I+D en SpazioDati

All you need is lo^H^HDATA

- Una (ex)start-up, operativa desde 2012
- ~30 personas en 2 sedes en Italia (Trento y Pisa)
- Especialista en (big)data collection/integration/interpretation
- “Business intelligence”



Atoka: Construye y vendrán

El directorio más grande de información empresarial en Italia:

- IT: 6M de empresas, RU: 4,5M, UK: 4,1M
- Info freshness: 10m - 5d

Key value proposition:

Integrar información oficial *y no oficial* sobre las empresas

- [Demo time](#)



Atoka: Componentes clave



entity extraction, text similarity, text classification, language detection and sentiment analysis.

Several languages (de | en | es | fr | it | pt). Works well with short texts.

- Which companies are mentioned in this text?
- Do text A and B contain similar entities?
 - Is this text about politics? science?
 - Is this text Italian? English?
 - Are they good reviews?

try it: <https://dandelion.eu/semantic-text/entity-extraction-demo/>

Atoka: Componentes clave

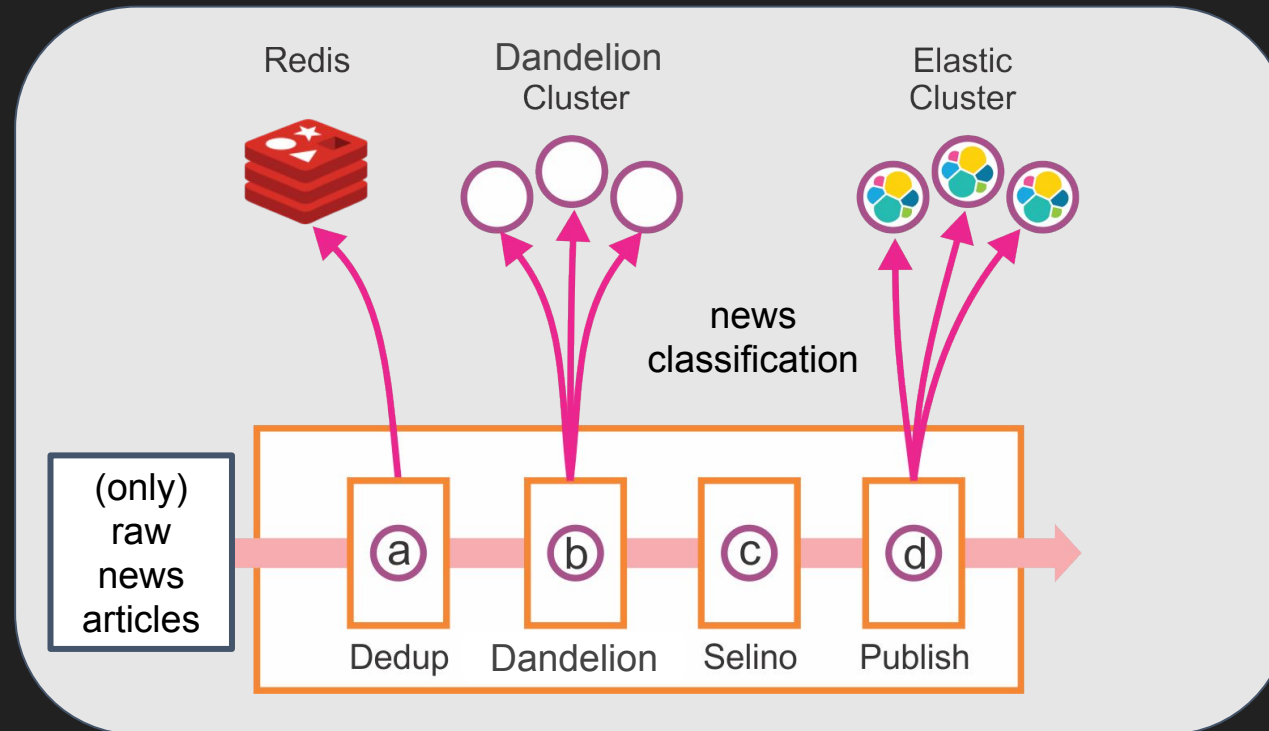
Sedano

A system for retrieving, annotating and indexing a continuous stream of business-related news.

- Which are the news where *a particular company is mentioned?*
- Which are the interesting events regarding *a specific company*, like mergers and acquisitions or product launches?

Atoka: Componentes clave

Sedano pipeline



Atoka: Componentes clave

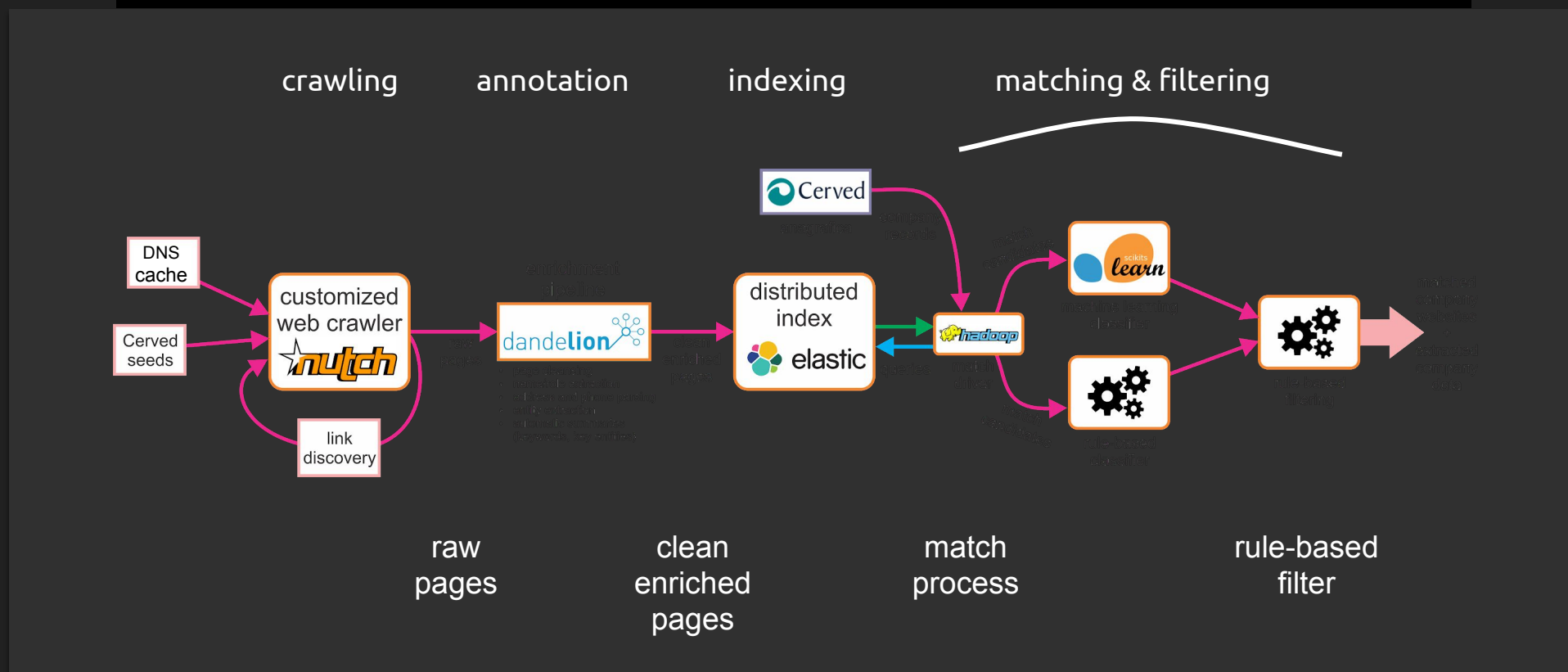
CWC: Corporate Web Crawler

An eclectic collection of crawlers to understand and extract structured information from the unstructured the corporate web

- Describe what company X is about (keywords)
- I want to contact people in charge of business development within company Y
 - Keywords What is the ownership distribution of company Z?

Atoka: Componentes clave

CWC: pipeline



Atoka: Componentes clave

CWC en números

- 198M de links procesados mensualmente (15M websites)
 - 60% fuera del grafo (real world is messy)
 - 30% links que apuntan al exterior
 - 10% que permanecen adentro
 - **1.5M** links entre compañías distintas
- **300k** compañías con cuentas en redes sociales
- **100k** compañías con info sobre plataforma e-commerce en uso (ej., Magento, WooCommerce, Opencart)

Atoka: el modelo de negocios

Market size: **EUR 200M** (Sales Intelligence - Lead Generation: EU, North America)

IPR: SpazioDati

Canales de negocios

1. Direct sales (subscription model)
2. Specialisations
 - Custom data
 - Data analysis on customer request
3. Alternative distribution channels through integration into other solutions
4. Several packaging options

Los money makers son: 2, 3, 4, 1 (en ese orden)

Atoka: otras integraciones

Debido a la **masa crítica** de datos, muchas oportunidades de integración de más y más fuentes de datos:

- Ofertas de trabajo
- Información contable
- Accionistas
- Indicadores especializados (ej.: Score de Innovatividad)
- Propiedad inmobiliaria
- Información territorial

Atoka: take home message

- Primero la necesidad clave, luego el resto de los features
- Llegar primero a un mercado vale muchísimo
- Muchas veces cambiar el paquete hace que las cosas comiencen a funcionar
- Cuantos más datos tengas, más oportunidades de integrar otros datos (círculo virtuoso)
- All you need is data: open data, official data, bring-your-own-data

Proyecto SmartTraffic

- Proyecto convocatoria Conacyt: colaboración UC Asunción y UniTrento
- Soluciones tecnológicas para la movilidad urbana en Asunción
- Tecnologías en el contexto de SmartCities:
 - Sistemas Colaborativos Adaptativos CAS
 - Mobile Crowdsensing
 - Participación ciudadana
- Tenemos una encuesta en la que vos podés ayudarnos:

<https://t.co/Bpy5RTGeY6>

