

Fusión de Información, Minería de Redes Sociales y NLP

Daniel Campos D.
Universidad Andrés Bello
dcampos@unab.cl



¿Qué queremos lograr?

- Apoyar acciones de inteligencia a través del desarrollo de una tecnología para el análisis automatizado de medios sociales, incluyendo redes sociales, de modo tal que los resultados de dicho análisis puedan ser visualizados en el plano espacio-temporal para ser utilizados por analistas humanos.



Visión del Producto

- Una aplicación que permita, al encargado de contacto con los clientes de la empresa, simplificar el monitoreo de la actividad de redes sociales permitiéndole concentrarse en las actividades de marketing general y directo y en el contacto con clientes de un modo más eficiente y con apoyo automatizado.
- Una aplicación con una arquitectura modular y evolutiva, a la cual puedan ir agregándosele nuevas capacidades a medida que la investigación avanza y éstas son desarrolladas y disponibilizadas a los clientes.
- La solución debe poder ofrecerse como servicio o como producto con soporte.



Desafíos principales

- Extracción de actividad general y de usuarios específicos desde redes sociales
- Identificación de focos de convergencia de opiniones e intereses.
- Análisis situacional (oportunidades y amenazas)
- Visualización de información

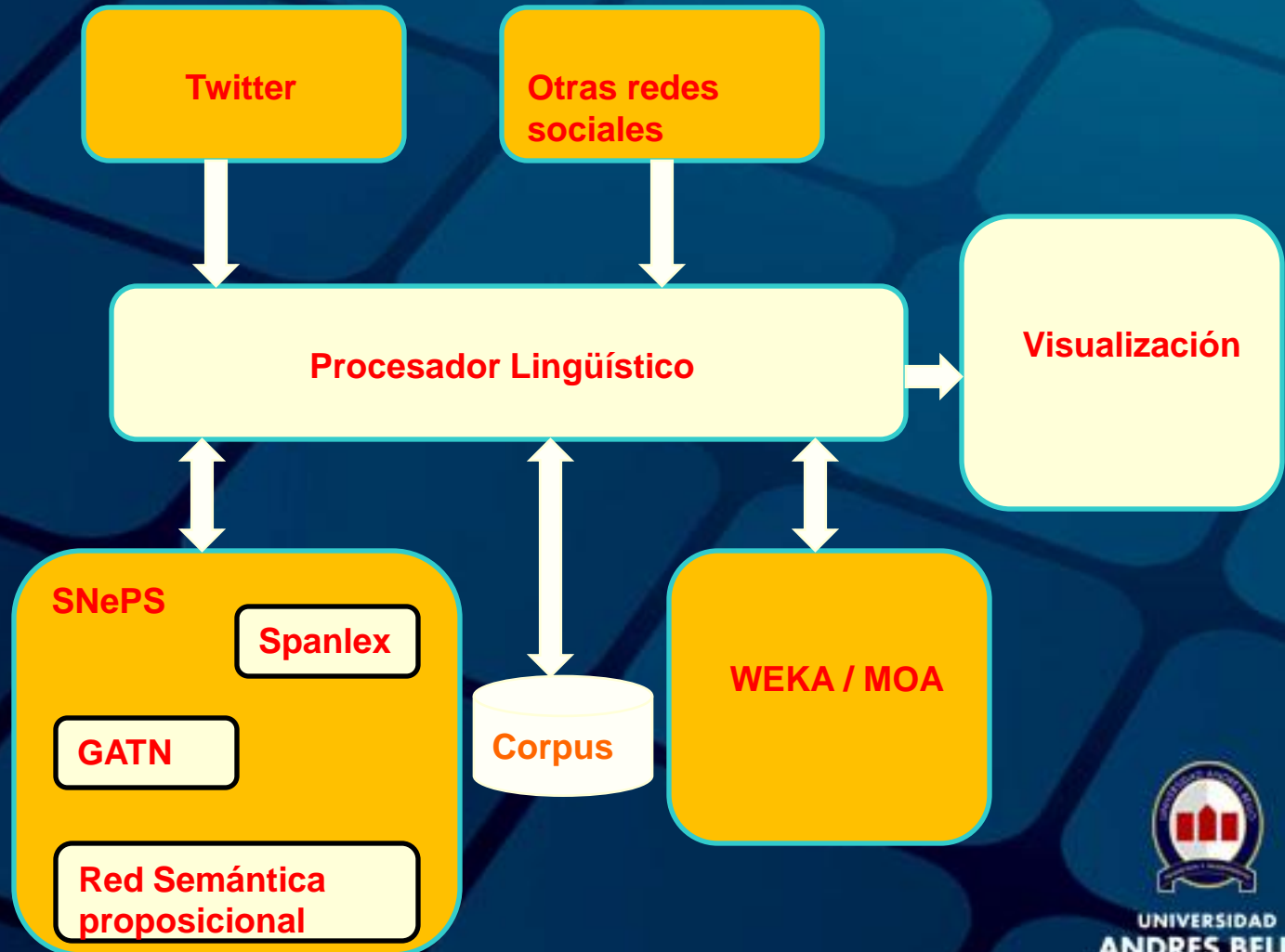


Focos de investigación

- Aplicación de técnicas procesamiento lingüístico para la extracción de evidencias a partir de contextos no estructurados (ej: intercambios de información en redes sociales)
- Aplicación de técnicas de minería de datos sobre información generada en medios y redes sociales (stream mining) para hacer emerger de dichas evidencias información considerada relevante (minería de eventos y sentimientos).
- Aplicación de técnicas de visualización en la representación de los resultados y minería de datos visual (visual mining)



Plataforma de investigación



Herramienta de investigación: TwitterSneps

The screenshot displays the Twitter-SNePS web application interface. At the top, there is a navigation bar with 'File' and 'Help' menus. Below this, a search form is visible with the following fields: 'Usuario:' containing 'dcamposd', 'Password:' with masked characters, and 'Buscar:' containing 'temblor'. A 'Login' button is positioned to the right of the search field. Below the search form, a horizontal menu contains several tabs: 'Tweets', 'Lexicon', 'GATN', 'WWFs', 'Clusters', 'Oportunidades/Amenazas', 'Búsqueda reglas GATN', and 'Stats'. The main content area is divided into three vertical columns, each displaying a list of search results for the term 'temblor'. Each result consists of a phrase followed by a count in parentheses. The first column lists results such as 'temblor (99)', 'de (61)', 'en (54)', 'el (42)', 'un (36)', 'rt (35)', 'no (31)', 'y (31)', 'al (28)', 'a (26)', 'que (26)', 'hubo (22)', 'fue (20)', 'me (19)', '#temblor (14)', 'mi (14)', 'todos (14)', 'la (13)', 'mismo (13)', 'o (12)', '40 (11)', 'cuando (11)', 'del (11)', 'km (11)', 'pendejo, (11)', 'por (11)', 'saltamos (11)', 'tiempo (11)', 'cerca (10)', 'como (10)', 'las (10)', 'richter (10)', and '- (10)'. The second column lists results such as 'temblor en (21)', 'el temblor (16)', 'hubo temblor (16)', 'un temblor (14)', 'al mismo (12)', 'temblor de (12)', 'mismo tiempo (11)', 'no peñejo, (11)', 'pendejo, saltamos (11)', 'saltamos todos (11)', 'todos al (11)', '40 km (10)', '- hubo (10)', '?', '- (9)', '@diariojoven: - (9)', 'en el (9)', 'rt @diariojoven: (9)', 'tiempo .xd (9)', 'cerca de (8)', 'temblor y (8)', 'a las (6)', 'del temblor (6)', 'en #salta (6)', 'km al (6)', 'al o (5)', 'de santa (5)', 'en la (5)', 'o de (5)', 'que no (5)', '#temblor #tucuman (4)', '15.37 horas, (4)', and '3.5 richter (4)'. The third column lists results such as 'al mismo tiempo (11)', 'no peñejo, saltamos (11)', 'pendejo, saltamos todos (11)', 'saltamos todos al (11)', 'todos al mismo (11)', '- hubo temblor (10)', '?', '- no (9)', '@diariojoven: - hubo (9)', 'mismo tiempo .xd (9)', 'rt @diariojoven: - (9)', 'temblor?', '- (9)', 'temblor en #salta (6)', '40 km al (5)', 'al o de (5)', 'km al o (5)', '15.37 horas, con (4)', '3.5 richter fue (4)', '40 km de (4)', 'a 40 km (4)', 'a las 15.37 (4)', 'con epicentro cerca (4)', 'cuando pase el (4)', 'de 3.5 richter (4)', 'epicentro cerca de (4)', 'fue temblor a (4)', 'horas, con epicentro (4)', 'km de profundidad (4)', 'las 15.37 horas, (4)', 'o de santa (4)', 'richter fue temblor (4)', and 'temblor a las (4)'. At the bottom of the interface, there is a 'Mensaje' section with the text 'Conectado a Twitter...' and a 'New tweet' button. A tooltip is visible over the 'New tweet' button, displaying 'From: "S_Schwartzmann"'. The background of the application window is a light beige color.



¿Con qué apoyo contamos?

- Tesis de Magister en Ingeniería Informática, estudiantes de pregrado de Ingeniería Civil en Computación e Informática y de Ingeniería en Gestión Informática.
 - Minería de sentimientos
 - Representación de situaciones como triples RDF
- Investigación previa en NLP :
 - Desarrollo de analizador morfológico en Español para SNePS (spanlex)
 - Modelamiento de relaciones temporales en español
 - Desarrollo de gramática de transición aumentada de oraciones en Español
- Contacto con equipos de investigación en IA en USA (SNeRG)
- Contacto con centros de investigación en Data/Information Fusion en USA (CMIF)



Posibilidades de colaboración

- Buscamos interesados en la aplicación de técnicas lingüísticas a la minería de redes sociales para el desarrollo de investigación y soluciones para empresas.
- Necesitamos apoyo en guía y comisiones de tesis.
- Nos interesa establecer convenios de intercambio académico y estudiantil.

Contactar:

Daniel Campos Dominguez

Email: dcampos@unab.cl

Fono: (56 2) 661 5889

Skype: danielcamposdominguez



Fusión de Información, Minería de Redes Sociales y NLP

Daniel Campos D.
Universidad Andrés Bello
dcampos@unab.cl

