

Gestión del Conocimiento: Sistema Colaborativo para la Innovación Local

Lic. Damián Barry¹, Lic. Santiago Raynoldi²

¹Depto. de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco. Puerto Madryn, Argentina +54 280-4472885 – Int. 117.

²Organización Industrial I - Universidad Tecnológica Nacional – Facultad Regional Chubut. Puerto Madryn, Argentina +54 280 154405205.

demian.barry@gmail.com, sraynoldi@gmail.com

Resumen. La producción y obtención de información ha pasado a ser uno de los grandes activos de las organizaciones actuales. En este sentido el desarrollo y estudio de la generación, administración, explotación, clasificación e interpretación de información para la toma de decisiones se ha convertido en un desafío a nivel mundial.

Para poder abordarlo, no solo se requiere el soporte de científicos y tecnólogos en el área de la informática, sino además de la integración con investigadores y expertos de distintas áreas vinculadas con las actividades que se desean analizar y comprender, para generar verdadero valor a la información circundante.

El presente trabajo, basado los Sistemas de Innovación Regional de la República Argentina, propone el desarrollo de una Plataforma Tecnológica de Gestión de la Información que permita integrar a todos los actores y desarrollar un sistema colaborativo y virtuoso para la innovación regional, basada en la gestión del conocimiento y la democratización de la información.

La herramienta permitirá sistematizar el conocimiento existente en instituciones académicas, científicas, tecnológicas, públicas, y ONG's y ponerla a disposición de la comunidad en general para su uso y desarrollo.

La plataforma gestionará información pública de los actores del Sistema de Innovación Regional, donde aplicando técnicas de búsqueda y recuperación de información, de clasificación de información, construcción de perfiles, patrones y ontologías, permitan entender el comportamiento de las estructuras organizacionales de los actores, convirtiendo datos circundantes en información de interés para la región, aprovechando información de conocimiento tácita, codificándola para su utilización.

La solución se potencia mediante la utilización de técnicas de análisis de redes sociales como una herramienta para la gestión de información desde la perspectiva del comportamiento organizacional.

El trabajo describe el marco teórico en el que se desarrolla la plataforma, su funcionamiento y su potencial uso.

Palabras claves: Gestión del Conocimiento, Sistema de Innovación Regional, Sociedad de la Información, patrones, ontologías.

Abstract: The production and gathering information has become a great assets of existing organizations. research and development of generation, management, operation, classification and interpretation of information for decision-making has become a global challenge.

In order to deal not just the support of scientists and technologists are required in the area of computer science, but also to integration with researchers and experts in

various fields related to the activities to be analyzed and understood, to generate real value to the surrounding information.

This paper, based on developed papers of Regional Innovation Systems of Argentina, proposes the development of a Technology Platform Management and Knowledge Information allowing, as a tool, virtuously integrate local actors to achieve a collaborative system and virtuous for innovation management based on knowledge and information society.

This framework will allow systematize existing knowledge in academic, scientific, technological, public institutions and NGOs make it available to the wider community for their use and development.

The platform will allow to manage public information Regional Innovation System, applying techniques of search and information retrieval, information classification, profiles, patterns and ontologies, to allow them to understand the behavior of the organizational structures of the actors, converting data into information of interest to the region, taking advantage of tacit knowledge information, encoding for use.

The solution enhanced through to use techniques of social network analysis as a tool for information management from the perspective of organizational behavior.

1. Introducción

La producción y obtención de información ha pasado a ser uno de los grandes activos de las organizaciones, ya sean públicas, mixtas o privadas. En este sentido el desarrollo y estudio de la generación, administración, explotación, interpretación y clasificación de información se ha convertido en un desafío tecnológico y científico a nivel mundial.

Para poder abordarlo, no solo se requiere del soporte de científicos y tecnólogos en el área de la informática sino además de la integración con investigadores y expertos de distintas áreas vinculadas con las actividades que se desean analizar y comprender, donde a través de la conformación de equipos multidisciplinarios generen verdadero valor a la información circundante.

En este sentido Edgar Morin [Morin99] en su libro “El Método” dice que “La idea de cibernética– arte-ciencia del gobierno puede integrarse y transformarse en co-cibernética – arte-ciencia de pilotear conjuntamente, donde la comunicación ya no es útil del mando, sino una forma simbólica compleja de organización.”

“La información generativa y la información circulante pueden transformarse la una en la otra, pero la transformación de una información circulante o de señales de información generativa no es posible más que si se encuentra un aparato capaz de registrarla y tratarla.”

“Así la información solo puede nacer a partir de una interacción entre una organización generativa y una perturbación aleatoria al ruido. Ergo la información no puede desarrollarse más que a partir del ruido. Y desde luego, en el nacimiento de una información, siempre se necesita una **actitud organizacional de carácter neguentrópico** que se supere a si misma transformando el evento en novedad.”

La democratización de esta información tiene como objetivos estimular y gestionar flujos de conocimiento y tecnología entre universidades, instituciones de investigación, organizaciones, empresas y mercados; impulsando el desarrollo del sistema productivo regional.

El presente trabajo tiene como finalidad mostrar un conjunto de herramientas concretas que permita mejorar la gestión del conocimiento de grandes volúmenes de información provenientes de los repositorios y sistemas de información digital, de las organizaciones del Sistema de Innovación Local, complementando la misma con información pública de interés y generando lineamientos para su utilización.

2. Antecedentes

2.1 Cambio de paradigma: Globalización y sociedad de información

Podemos observar que estamos participando de un cambio de paradigma, de una transformación de los criterios básicos con los que comprendemos la realidad: Durante muchos siglos se conocieron discursos (religiosos, culturales, políticos y científicos) que pretendieron unificar; siempre se propuso una explicación que redujera lo que sucede a un solo principio: lo que Derrida llamó “logocentrismo” y, en palabras de Gilles Deleuze, se podría denominar “monológica”.

Con la lógica virtual, la lógica informática, la presencia de Internet, ya se está instalando como aquel criterio en el que se disuelve toda idea de centro.

Por lo tanto a partir de la obtención de esta información y aplicando técnicas de clasificación según criterios de expertos es que podemos transformar el ruido en verdad.

El acceso democrático a la información pública generada por personas y organizaciones, nos permite, a través de técnicas adecuadas, detectar patrones para su clasificación y realizar transformaciones concretas para la obtención de nueva información útil que permita retro-alimentar el aprendizaje en un proceso continuo neogentrópico.

En particular si analizamos la problemática dentro del marco del desarrollo socio-productivo respecto de la información como conocimiento, refuerza este concepto lo expuesto por Yoguel, Robert, Erbes y Borello en su trabajo “Capacidades cognitivas, tecnológicas y mercados” donde se expresa que: “... En este nuevo esquema el conocimiento presenta una característica distintiva respecto a otros factores de producción, su consumo aumenta la producción y no se agota al utilizarse [Cimoli y Correa, 2005, Yoguel 2000]. El consumo y circulación del conocimiento, que asimismo constituyen una fase importante de su producción, depende positivamente de la complejidad y articulación de las redes y del grado de competencias endógenas de los agentes involucrados y de las complementariedades que se generan en la interacción entre los agentes que las integran [Ocampo, 2005] ...”

Comienza a existir un creciente acuerdo acerca de que la innovación no constituye un fenómeno individual de firmas u organizaciones [Freeman et al 1991] y crecientemente se enfatiza en el carácter multi-direccional del proceso de aprendizaje, cada vez mas contextual y transdisciplinario con una fuerte interacción y complementariedad de los componentes tácitos y codificados [Nonaka y Takeuchi, 1995]. En ese proceso, comienzan a cobrar creciente importancia los mecanismos de intermediación y traducción que se manifiestan en la emergencia de las llamadas instituciones puente [Casalet, 2000a, 2000b]. ...”

2.2 El Hombre como generador de conocimiento

Según [Yoguel et al, 2009] “...en algunos estudios se examinan las posibilidades de crear ventajas competitivas a partir de la producción y el intercambio de conocimientos dentro de una trama productiva [Bisang y otros, 2004; Albornoz, Milesi y Yoguel, 2004; Yoguel, Novick y Milesi, 2003; Casalet, Cimoli y Yoguel, 2005]. Una de las configuraciones más virtuosas de red se vincula a la concentración geográfica y la generación de externalidades que promueven la cooperación y la eficiencia colectiva [Schmitz, 1995; Meyer-Stamer, 1998]. De esta manera, mientras algunas tramas tienden a ser menos eficaces [Humphrey y Schmitz, 2000; Poma, 2000], en otras se desarrollan procesos cooperativos de aprendizaje y difusión del conocimiento tácito [Freeman, 1988; Camagni, 1991; Nonaka y Takeuchi, 1995].”

La construcción y desarrollo de herramientas de gestión colaborativa, permiten mejorar las condiciones planteadas. Continuando con el trabajo de Yoguel donde define al Sistema de Innovación Local: “Entendemos por sistema local de innovación el espacio de interacción entre empresas y entre empresas e instituciones, en una ubicación geográfica común que incluye tanto las relaciones de competencia como de cooperación. Estos sistemas son heterogéneos y de muy diversos grados de complejidad, la que depende del número y las características de los actores y agentes que los componen y de las vinculaciones formales e informales entre ellos. De este modo,

cualquier sistema puede ubicarse en una serie de situaciones que van desde un nivel de máxima virtud —cuando hay procesos significativos de aprendizaje y creación de ventajas competitivas dinámicas— hasta el extremo opuesto en que ambos son casi inexistentes.”

La conectividad (Internet), intercambio de información, redes de información de todo tipo, interacción de las partes e inter-operabilidad, inteligencia colectiva, gestión del conocimiento, transformación de valores y saberes tácitos en codificados, simulación social y la construcción de nuevos espacios semánticos y sociales, serán el desafío tecnológico de los sistemas de información del futuro que garantizarán una mejor calidad de vida al hombre asegurando una mejor utilización de los recursos económicos, profesionales, logísticos, culturales y sociales disponibles dentro de cualquier tipo de organización política y garantizando el equilibrio entre todas las partes, tanto públicas como privadas, equilibrando sus contradicciones y potenciando sus cualidades. Obligando al uso racional de la información y a reducir la brecha sobre el conocimiento que existe actualmente.

Además, como se ha mencionado la gran proliferación de información pública generada en internet, ha provocado un fenómeno de identidad donde pareciera que la sociedad tiende a la uniformidad de criterios y comportamientos. Actualmente las comunidades ligadas a la información en Internet se están re-ordenando, generando un nuevo orden conceptual sobre las identidades y particularidades en sus comportamientos.

Esto es demostrado por la creciente influencia de las redes sociales, especialmente en los jóvenes que han cambiado sus hábitos de encuentro social. El incremento en la necesidad de generar estas particularidades formando grupos de interés con características geográficas, ideológicas y culturales comunes permiten identificar la necesidad de clasificar la información según los segmentos mencionados con la finalidad de segmentarla en ordenes de interés.

La segmentación de la información en áreas de interés ligados a los perfiles de cada individuo, permitirán a los distintos usuarios de la red encontrarse mejor con la información buscada.

Para ello es necesario contar con:

- Extracción de información pública de estructura heterogénea. Utilización de técnicas de Crawling e Informatrion Retrieval
- Segmentación de la información mediante etiquetas: clasificación, segmentación, generación automática de atributos.
- Ecuilizador de interés: perfil del usuario basado en pertenencia geográfica, identidad cultural y hábitos.
- Motor de búsqueda que jerarquice la información analizando la clasificación de la misma y los hábitos del usuario.

Desarrollar una herramienta y facilitar al SIL una plataforma para la creación de información que contemple mecanismos de captación y traducción de conocimientos tácitos en las organizaciones cobra importancia y genera nuevos datos para su utilización que fortalecen el sistema. Socializar esta información y complementarla con la existente potencia el proceso de desarrollo, innovación y crecimiento del sistema productivo y de los actores institucionales involucrados.

3. Desarrollo "La Plataforma"

¿Por qué crear una Plataforma Tecnológica? Teniendo en cuenta el pensamiento de Jorge A. Sábato, "...el conocimiento técnico y científico es una mercancía que como tal se produce, se distribuye, se vende, se compra, se exporta, se importa, se intercambia, y es por eso que hay que manejar el problema de producción de tecnología en el país no ya como derivado espontáneo de la producción de conocimientos, sino como un objetivo específico del sistema económico...". Esto destaca como un proceso político, la acción de insertar la ciencia y la tecnología en la trama misma del desarrollo que significa saber dónde y cómo innovar. Así, a través de la acción coordinada de la infraestructura del estado, de la infraestructura científico-tecnológica y de la estructura productiva, puede lograrse el desarrollo.

De lo expresado en el Plan Nacional ARGENTINA INNOVADORA 2020, surge que la dificultad de transferir los conocimientos generados en el sistema de Ciencia y Tecnología al sector productivo ha sido uno de los puntos de preocupación de las políticas de Ciencia, Tecnología e Innovación en los últimos años. Tal como se puntualiza previamente, en los últimos quince años y en un número importante de diagnósticos en el tema, el Sistema Nacional de Ciencia, Tecnología e Innovación se caracteriza por una baja articulación público-privada y entre privados.

En materia de innovación productiva el país sigue mostrando una actividad limitada, y se deben extremar los esfuerzos en marcha para vincular los conocimientos generados en el ámbito de la Ciencia y la Tecnología con las necesidades del sector productivo, así como para incentivar en las empresas comportamientos innovadores. En este sentido, las empresas suelen desarrollar esfuerzos individuales de innovación que generalmente se traducen en la obtención de innovaciones incrementales y su participación en el gasto en I+D+i es reducido. Este escenario es bastante concluyente, en consecuencia, con respecto a que el impulso a la innovación se realiza fundamentalmente “desde arriba”, ya que, en general, no es una inquietud que surja de las propias empresas. Esto explica que, en el pasado reciente, las políticas de Ciencia, Tecnología e Innovación hayan mostrado una preocupación recurrente por promover una mayor transferencia de los conocimientos generados en las instituciones de Ciencia y Tecnología hacia el sector productivo.

En este marco, el Ministerio Nacional de Ciencia, Tecnología e Innovación Productiva (MINCyT) pone un fuerte énfasis a través del Plan en la articulación público privada, buscando generar los mecanismos de interrelaciones más fuertes y dinámicas que permitan la circulación del conocimiento científico tecnológico en el ámbito productivo. La idea fuerza que subyace es que el sistema público y el privado actúan con lógicas diferentes pero complementarias. El primero busca optimizar la generación de información y conocimiento con los recursos disponibles, mientras que el segundo busca tomar y aprovechar la información para convertirla en un producto o servicio que llegue a la población. La conexión entre ambas lógicas resulta, por ende, fundamental para que el quehacer científico y tecnológico se traduzca en mejoras productivas y sociales.

Por lo tanto la lógica de derrame de conocimiento debe ser reemplazada por una lógica dinámica de articulación, interacción e integración entre los productores de conocimiento y los receptores del mismo. Esta tarea se refleja en la propuesta de un Sistema de Innovación Local donde, desde la estructura de Estado+Ciencia+Producción, se vea dinamizada por un círculo virtuoso entre los actores del Sistema.

En este sentido entendemos que la propuesta de creación de una Plataforma Tecnológica en Gestión de la Información y el Conocimiento, debe estar fuertemente acompañada de contenido y acciones que logren que meros enunciados de ideas se concreten y transformen en realidades de innovación y producción. Es necesario comprender en este sentido que el enunciado de articulación tampoco es suficiente en sí mismo, no alcanzará con sentarnos en una misma mesa, sino en establecer acciones que permitan dinamizar y lograr los objetivos de innovación como fin último.

Por lo tanto el sentido de los ejes estratégicos propuestos en el presente documento tienden a garantizar y viabilizar una verdadera articulación de Estado+Ciencia+Producción y las acciones necesarias para lograr que se produzca la innovación deseada.

Lograr el equilibrio de los actores del Sistema de Innovación será el desafío de la Plataforma tecnológica como organización dinámica, pues estas fuerzas por naturaleza tienden a diverger. Con responsabilidad y participación, los actores del sistema deberán encontrar los mecanismos y acciones para lograr el equilibrio deseado y será por lo tanto misión fundamental de las instituciones a través de sus impulsores y traductores tecnológicos, garantizar este equilibrio.

3.1 Herramientas desarrolladas para la plataforma

La herramienta ha sido desarrollada en el Centro de Investigación de la Universidad Nacional de la Patagonia San Juan Bosco, a través de un proyecto de investigación Dirigido por el Lic. Demian Barry, durante el cual, se han programado los códigos para la construcción del motor de extracción de información, motor de etiquetas y el ecualizador de interés, módulos que se describen en el presente trabajo. Además, se contempla la estrategia de implementación

3.1.1 Construcción de un motor de extracción de información

Para la recuperación de contenido público de internet se desarrolló un motor de extracción (Crawler) denominado dentro del proyecto como Zcrawler, el mismo se compone de un lenguaje de extracción que le permite a los usuarios mediante una especificación comenzar a extraer información según múltiples criterios, permitiendo la clasificación de esta información según a quien se está evaluando.

La estructura básica de extracción permite definir una zona geográfica vinculada con la información a extraer, esta decisión se enmarca en la necesidad de georreferenciación de contenidos. Luego se especifica la fuente a extraer, indicando al motor de extracción el componente específico de extracción a utilizar. Además de la georreferenciación se permiten utilizar otros clasificadores mediante la utilización de tags. Finalmente el motor de extracción permite configurar la frecuencia de extracción.

La herramienta de extracción contempla un conjunto de utilitarios que permiten a quien realiza las configuraciones de extracción realizar análisis previos de la información a extraer y simular resultados de extracción para evaluar si es lo que se desea.

Actualmente se han desarrollado los servicios de extracción para facebook, twitter, RSS estándar y además se han adaptado parser específicos para blogs y diarios en línea que no cuentan con una implementación estándar de RSS. En particular se implementaron conectores para los diarios Jornada y Chubut, ambos de la provincia de Chubut.

Adicionalmente para los grandes medios nacionales se adaptaron extractores ya que sus definiciones de RSS no accedían ni a los comentarios ni a las referencias (links) externas ni a las imágenes.

A partir de poder recuperar imágenes de las noticias y publicaciones extraídas se mejoró la herramienta permitiendo realizar un tratamiento común a todo el material multimedia asociado a una publicación realizando una gestión uniforme tanto para facebook, twitter y las páginas sindicadas.

3.1.2 Clasificación de la información.

El etiquetado automático (autotagging) se incorpora al proyecto como un modo de colaborar con la carga de contenido. El objetivo fundamental es que tanto en la carga manual de contenidos como en su incorporación automática mediante medios de sindicación (Atom, RSS, etc.) o extracción de redes sociales, el editor cuente con opciones de tags sugeridas por el sistema, teniendo en cuenta aspectos de espacios semánticos que fueron previamente determinados por expertos en distintas categorías de noticias.

La versión original sugiere tags en función de las palabras de un contenido. Esto no es óptimo porque las palabras no definen la temática de un contenido. Muchas veces son frases o construcciones complejas las que cumplen este cometido. Para esta tarea se han utilizado características avanzadas de la herramienta de gestión de contenidos utilizada la herramienta base utilizada Apache-Solr.

Al igual que sucede con la web semántica donde es necesario generar atributos en la información para poder contextualizarla, nos era imprescindible generar mecanismos que permitieran clasificar la información.

El recurso natural dentro de los Gestores de contenidos (CMS) es la clasificación de los artículos en secciones y categorías. Dicha clasificación, claramente jerárquica, es sumamente rígida a la hora de necesitar una visión multidimensional de la información.

Para solucionar esto se recurrió a técnicas de data-warehousing utilizando un modelo estrella dinámico para segmentar la información. Donde cada artículo es clasificado, mediante etiquetas, dentro de estructuras de grupos jerárquicos. De esta forma se logra que la información no solo esté etiquetada sino que también permite ser agrupada jerárquicamente.

A modo de ejemplo y tomando como grupo de etiquetas a la localización geográfica podemos contar con la siguiente relación jerárquica:

Comuna →
Localidad o Ciudad →
Partido o Departamento →
Provincia o Estado →
País →

De forma similar se podría contar con información de interés sobre determinadas áreas, como por ejemplo deportes:

Nombre de equipo de fútbol →
Fútbol →
Deportes de equipo →
Deportes →
Deportes de balón
Deportes →

Como podemos observar además se pueden compartir distintos niveles de jerarquía mediante una etiqueta como el del nombre de nuestro equipo de fútbol favorito.

3.1.2.1 Motor de etiquetas.

El motor de etiquetas hace uso de las funcionalidades de Apache Lucene, analizando los contenidos a etiquetar y se extraen, mediante técnicas de information retrieval, los términos más significativos del documento.

Primeramente se reducen los términos de significado idéntico o muy cercano reemplazándolos por un sinónimo en común. A continuación se procede con la lematización o stemming, proceso que reduce las palabras a su parte esencial (lema), quitando conjugaciones, prefijos y sufijos:

automóvil, automotor → auto

De esta forma, logramos un conjunto uniforme de lemas con los que se puede construir un ranking de ocurrencias y fijar un umbral considerando aquellos que lo sobrepasan como posibles etiquetas candidatas.

Una vez identificados estos términos, aquellos que existan en la base de datos de etiquetas (literalmente o por sinónimos) son sugeridos al autor como etiquetas candidatas.

Cuando se ubican dos o más etiquetas candidatas dentro de la misma jerarquía, se descartan aquellas con un grado de ocurrencia bajo en relación a las demás con el fin de no entorpecer la clasificación cuando aparecen sólo circunstancialmente en el texto. Se evita de esta forma que si se habla de fútbol y circunstancialmente se lo contrapone con básquet, el motor sugiera "básquet" como etiqueta cuando el artículo sigue tratándose de un deporte y no del otro.

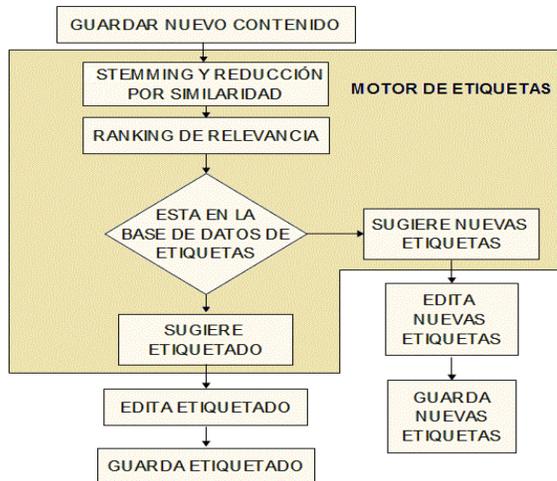
Asimismo asignamos mayor importancia a las etiquetas candidatas que aparezcan en el título del texto por sobre las que figuren en el cuerpo o aún en epígrafes y descripciones secundarias.

Este mecanismo asegura que la clasificación mediante etiquetas sea relativamente independiente del criterio del autor, dado que es el motor el que indica las etiquetas que pueden corresponder al texto indicado, sin impedir que en casos particulares dicha clasificación sea modificada.

Otro proceso importante que surge como producto colateral del motor de etiquetas son las sugerencias de nuevas etiquetas, lo que permite indagar nuevas clasificaciones y jerarquías. Esto es vital, ya que la dinámica de los portales y de la generación de información y la idiosincrasia de cada lugar requerirán de un eficiente método que permita detectar nuevas estructuras jerárquicas de clasificación.

El motor de etiquetas extrae aquellos términos de gran relevancia dentro del texto y que no han sido ubicados en la base de datos de etiquetas, y sugiere su inclusión dentro de la jerarquía correspondiente.

El funcionamiento del motor puede diagramarse de la siguiente manera:



3.1.3 Ecuador de interés.

El ecuador de interés es la definición de los perfiles por parte de cada usuario de la comunidad con la finalidad de garantizar que la información recuperada dentro de la plataforma está especializada y jerarquizada según los intereses particulares de los habitantes de una región.

Para ello es importante brindar una herramienta de administración de usuarios y configuración de perfiles que permita a cada usuario tanto configurar un ecuador de interés basado en su zona geográfica, como guardar distintos ecuadores para otras zonas.

El usuario registrado tendrá acceso a un ecuador de interés a través del cual podrá seleccionar y filtrar los contenidos según sus preferencias. Haciendo un portal ampliamente adaptable y configurable por cada usuario en particular.

El editor de preferencias permite indicar para cada estructura jerárquica de etiquetas, bandas de preferencias donde se establece el peso deseado para cada zona en particular.

3.1.4 Articulación con el SIL

Tenemos el paradigma de que la información posee un valor estratégico y competitivo para el que la posee, y generalmente se tiende a no socializarla, individualmente y desde las organizaciones, sean estas públicas o privadas, impidiendo que su utilidad no permita el beneficio para el que la necesita. Ahora, teniendo en cuenta la tendencia actual de lograr mejores prácticas y calidad de vida a través del uso de información, tener a disposición una plataforma con acceso a información pública, que pueda administrarla y gestionarla permitirá a un mayor número de usuarios de acceder a sus beneficios. El desafío es lograr que las organizaciones etiqueten la información que se genera dentro de ellas como pública, que se encuentren en reservorios con acceso y que puedan ser explorados por la plataforma tecnológica y sus herramientas de búsqueda y clasificación. En ese sentido se está trabajando en nuestra ciudad a través de Parque Tecnológico de Puerto Madryn.

3. Resultados

Los resultados de las actividades realizadas se pueden dividir en:

- Desarrollo y prueba de la herramienta
- Articulación y sensibilización entre los actores del Sistema de Innovación Local

El equipo investigador ha desarrollado, integrado, configurado y realizando una prueba de la herramienta "Zcrawler", buscando en repositorios de información pública, obteniendo información durante 6 meses, que arrojaron como resultado 1.2 millones de documentos. Estos documentos se

clasificaron y georeferenciaron de acuerdo a criterios pre determinados por el equipo, para su utilización.

Actualmente, “Zcrawler” está disponible para su uso y a disposición de las instituciones locales.

Ahora, desde el punto de vista de su utilización de acuerdo a lo expuesto, los resultados son más difusos, porque se requiere la aceptación de los actores del Sistema de Innovación Local para compartir información que puede ser pública, pero que no se encuentra disponible ni alcanzada por herramientas como buscadores, meta buscadores u otras. Se debe afianzar dentro de las organizaciones el concepto de socialización de la información y del conocimiento y se debe generar el hábito de guardar esta información en repositorios públicos para poder acceder a ella. A modo de ejemplo se puede mencionar la disposición como elemento público de trabajos académicos, de investigación, informes técnicos, actas, estudios, u otro documento de interés que contenga información en temas diversos, generados de acciones desarrolladas en el seno de las organizaciones y que son de su propiedad.

Actualmente, fuera del alcance de este trabajo, pero de manera vinculada, se está desarrollando en el ámbito de nuestra ciudad una Oficina de Vigilancia Tecnológica, promocionada por el Programa VinTec del MYNCYT, que la integran las instituciones públicas, académicas, de ciencia y tecnología y empresariales de la ciudad. En ese ámbito se está discutiendo la utilización de herramientas para la búsqueda y análisis de información que permita la innovación en el Sistema Productivo Local. “Zcrawler” iniciará es pondrá a disposición de la OVT para su utilización al servicio de las empresas e instituciones locales.

Además, se prevé utilizarla para generar y codificar información existentes en sistemas cerrados (empresas) que permita vigilar los que ocurre en su red de información y emitir informes al respecto.

Las actividades de sensibilización realizadas hasta la actualidad, la disposición de la herramienta que facilita la tarea de crauer, la conformación de una oficina que busque y procese información transformándola en otra información útil o adecuada al momento en que se la necesita, va generando confianza y permite que todo el sistema se desarrolle.

4. Conclusiones

A modo de conclusión podemos decir que el proceso de extracción de información y su correcta clasificación es el primer paso analítico para el procesamiento de grandes volúmenes de información. Particularmente las herramientas de extracción y clasificación cumplen la misión de recolectar información de interés tanto para usuarios que requieren encontrar información según sus preferencias y ubicación geográfica y no por si la noticia tiene más impresiones que otra, premiando en este sentido la centralidad de la información.

La gran cantidad de información recuperada puede ser utilizada para el reconocimiento de patrones sociales que permitan tanto la identificación de “rumores” como la sensibilidad sobre ciertos aspectos y temas de sensibilidad social.

Aplicando técnicas de Análisis de Redes Sociales podemos identificar geográficamente grupos de personas con intereses comunes y vinculadores que permiten unir distintos clusters. Todo esto analizado especialmente desde la perspectiva geográfica, potenciando a la herramienta en una herramienta avanzada de vigilancia tecnológica o ser usada como herramienta de marketing político o clima social. Por supuesto todos estos aspectos requieren de la conformación de equipos multidisciplinarios integrando técnicas avanzadas de análisis de información con el reconocimiento de patrones específicos según ejes temáticos concretos evaluados por expertos en cada área.

Claramente la indexación y clasificación de información mediante etiquetas permite analizar la misma desde un punto de vista espacial, aportando en la resolución de las búsquedas una potencialidad inexistente en la actualidad a un bajo costo computacional.

El trabajo aquí presentado permite confirmar experiencia ya que ha sido sometido a importantes pruebas de stress tanto de volumen como de calidad de los resultados esperados en la evaluación de la clasificación de información y los intereses particulares de cada usuario.

Por último, concluimos que debemos socializar esta herramienta para fomentar la democratización de la información con el objeto de mejorar y dar impulso y fomentar la innovación en el sector productivo local, apoyado . Trabajar en ese sentido, articulando y generando consenso es el eje de trabajo a seguir.

5. Referencias

1. Apache Solr, <http://lucene.apache.org/solr/>
2. Borello, J.; Milesi, D.; Novick, M.; Roitter Sonia; Yoguel, G. (2003); Las nuevas tecnologías de información y comunicación en la industria argentina: difusión, uso y percepciones a partir de una encuesta realizada en la región metropolitana de Buenos Aires; en: Nuevas tecnologías de información y comunicación. Los límites en la economía del conocimiento; Boscherini, F.; Novick, M.; Yoguel, G. (comps.); Buenos Aires; Miño y Dávila Editores.
3. Cal Henderson: "Building Scalable Web Sites", O'Reilly Media, 2006
4. Damián Barry, Juan Manuel Cortez, Francisco Páez: Construcción de un Ecuilibrador de Interés Mediante el uso de Lucene-Solr, IEEE Intercon 2010.
5. Damián Barry, Ignacio Aita, Francisco Páez: Distributed Search on Large NoSQL Databases, PDPTA2011.
6. Erik Hatcher, Otis Gospodnetić. "Lucene in Action", 2nd. ed, Manning Publications Co. 2004.
7. Gabriel Yoguel, Verónica Robert, Analía Erbes y José Borello. Capacidades cognitivas, tecnologías y mercados: de las firmas aisladas a las redes de conocimiento. 2005.
8. Gabriel Yoguel, José A. Borello y Analía Erbes. Argentina: cómo estudiar y actuar sobre los sistemas locales de innovación. Revista CEPAL 99, Diciembre 2009
9. Gabriel Yoguel, Marta Novick, Darío Milesi, Sonia Roitter, José Borello. Información y conocimiento: la difusión de TICs en la industria manufacturera argentina. Reporte Interno del proyecto de investigación: "Las tecnologías de la información y la comunicación (TICs) en la Argentina: Origen, difusión y prospectiva".
10. Gustavo Lugones, Fernando Peirano y Diana Suarez. Estrategias empresariales de uso y aprovechamiento de las TICs. 33º JAIIO (Jornadas Argentinas de Informática e Investigación Operativa. Simposio sobre la Sociedad de la Información Córdoba, Argentina, 24 de Septiembre 2004.
11. Ignacio Katz, Retornar al pensamiento lógico. Revista Médicos Número, NEWSLETTER 82 / 24 de Marzo del 2003.
12. Jalfen, Luis J. Globalización y Lógica Virtual. Primera Edición, Ediciones Corregidor, 1998.
13. Javier Carrillo. Ciudades de Conocimiento: el estado del arte y el espacio de posibilidades. Transferencia, año 18, No. 69, enero de 2005, pp 26-28.
14. Morin, Edgar. El Método. Quinta Edición, Ediciones Cátedra, 1999.