

Una propuesta para la gestión colaborativa del conocimiento

Xavier Alamán, Ruth Cobos, Jaime Moreno

Departamento de Ingeniería Informática, Universidad Autónoma de Madrid

INTRODUCCIÓN

El Grupo de investigación sobre Herramientas Interactivas y Aplicaciones (GHIA) del departamento de Ingeniería Informática de la Escuela Técnica Superior de la UAM, en el marco del proyecto "KnowCat: Catalizador Automático de la Cristalización del Conocimiento" financiado por la Comunidad Autónoma de Madrid (07T/0027/1998), ha desarrollado un sistema distribuido llamado KnowCat que, sin necesidad de supervisión, permite la creación incremental de conocimiento estructurado de forma colaborativa. La motivación subyacente es permitir la generación de conocimiento colectivo de alta calidad como resultado de la interacción de los usuarios con dicho conocimiento.

Este sistema nos permite la construcción de "lugares Web" donde podemos encontrar conocimiento relevante y de calidad sobre un área o tema. Dichos lugares son "KnowCat sites" o nodos KnowCat a los que podemos acceder a través de la Web, mediante una dirección URL. El conocimiento de un nodo KnowCat está bajo un proceso de cristalización de conocimiento, mediante el cual tendremos en todo momento constancia de cuál es más relevante gracias a las opiniones y uso de éste por parte de los usuarios.

El sistema tiene dos áreas de aplicación: por un lado en el entorno docente, donde el objetivo es generar materiales educativos de alta calidad como resultado de la interacción colaborativa por parte de los estudiantes con los propios materiales; por otro lado en el entorno de un grupo de investigadores, cuyo objetivo es la creación y mantenimiento de una "memoria organizacional" donde compartir y evaluar el conocimiento relativo a su investigación conjunta.

En los siguientes apartados se encuentra: en primer lugar una descripción detallada del sistema KnowCat; en segundo lugar una síntesis de las experiencias realizadas con el sistema y los resultados obtenidos de éstas; finalmente se detallan las líneas de investigación abiertas en el área de investigación del sistema.

LAS CARACTERÍSTICAS DEL SISTEMA KNOWCAT

El nombre, KnowCat, es el acrónimo de "Knowledge Catalyser" o "catalizador de conocimiento", haciendo referencia a la propiedad principal que exhibe el sistema: la catalización del proceso de cristalización del conocimiento.

KnowCat trabaja con conocimiento explícito, es decir, el tipo de conocimiento que puede transmitirse a través de documentos, imágenes y otros elementos. Además KnowCat considera preferentemente el conocimiento que es estable en el tiempo. Ejemplos de este tipo de conocimiento son los que podemos encontrar en enciclopedias o libros de referencia.

El sistema organiza el conocimiento en un árbol jerárquico al que llamaremos árbol de conocimiento. La raíz del árbol es el tema principal del área de conocimiento a tratar. Cada tema o nodo del árbol tiene dos tipos de elementos asociados: un conjunto de descripciones del tema correspondiente y un conjunto de refinamientos del tema. El primero es un conjunto de *documentos* o unidades atómicas de conocimiento alternativas entre sí, las cuales tienen por objetivo dar una descripción del tema sobre el que tratan lo más acertada posible, es decir, son candidatos a describir el tema en cuestión. Un refinamiento, en cambio, es una lista de temas candidata a definir la descomposición de un tema concreto en sus subtemas principales. Dichos subtemas, naturalmente, son a su vez nodos pertenecientes al árbol de conocimiento.

Aunque, la unidad atómica de conocimiento es el documento, resulta de utilidad el tener la opción de poder acceder a partes de éste, y de esta manera permitir, por ejemplo, modificar

parte del documento, o añadirle algo nuevo en una parte determinada. Aparecen entonces las *anotaciones a documentos*. Las anotaciones están pensadas para permitir a un usuario colaborar aportando nuevas ideas sobre el documento de otro, es decir, una anotación a un documento es un comentario, aclaración, crítica o sugerencia sobre el conocimiento aportado por otro autor en forma de documento.

Sin embargo, cuando un usuario desea cambiar todo o parte de un documento aportado al sistema (bien sea añadiendo nuevo conocimiento, o modificando o quitando alguna parte del documento original) lo que está realizando es una versión del documento completo. Así, el "versionado" de un documento se entiende como la aportación de una revisión de una unidad de conocimiento del tipo documento por parte del mismo autor de la primera. Las revisiones propuestas por el autor de un documento contendrán las modificaciones que éste crea convenientes para mejorar la calidad de su documento, considerando o no las posibles anotaciones que recibiera su documento.

Cada una de las unidades de conocimiento comentadas están bajo su "propio" proceso de cristalización de conocimiento, en cada caso el proceso se adapta al tipo de interacciones que cada elemento de conocimiento es susceptible de recibir por parte de los usuarios. El proceso de cristalización sobre el conjunto de documentos o descripciones de un tema y el conjunto de refinamientos de un tema dado proporciona el elemento cristalizado (dominante), es decir, el elemento más aceptado en un momento dado. Cualquier otra descripción o cualquier otro refinamiento presentes serán considerados "candidatos" en pugna por obtener la dominancia a costa de los actuales. En el caso de las anotaciones, su proceso de cristalización es menos "exigente", ya que permite la supervivencia en el sistema de las anotaciones más aceptadas como candidatas a modificar el documento al que acompañan. En todos los casos, los elementos que no obtienen suficiente éxito al cabo de un tiempo son eliminados de la lista de candidatos (de su categoría).

El proceso de cristalización de los átomos de conocimiento o documentos se basa en el cálculo y evolución de lo que se denomina "grado de aceptación", valor que mide "cuánta aceptación ha recibido un documento por parte de los usuarios". En dicho cálculo intervienen todas las opiniones que recibe en forma de interacciones de los usuarios. De esta manera influyen en el cálculo tanto las opiniones explícitas de los usuarios mediante sus votos y anotaciones al documento (interacciones que modifican lo que se denomina la "aceptación explícita" del documento) como las opiniones implícitas de los usuarios a través de sus consultas o accesos a éste (estas últimas interacciones modifican lo que se denomina "aceptación implícita" del documento).

La evolución de la aceptación es lo que nos lleva al concepto de cristalización: un documento que es aceptado de forma constante en el tiempo es un candidato a cristalizar. En cambio, un documento con una evolución de su aceptación no constante, incluso negativa, es un candidato a ser eliminado del sistema. Un documento candidato a cristalizar lo hará cuando además de cumplir la condición anterior, su grado de aceptación esté por encima de un umbral fijado a priori. Pero un documento que cristaliza, también puede "descristalizar", es decir, dejar de ser el más aceptado, esto ocurre si durante un periodo de tiempo su grado de aceptación es menor a otro umbral fijo (menos al anterior) y por lo tanto la evolución de dicho valor es descendente.

Sin embargo, no sólo es importante el número de opiniones que reciben los elementos de conocimiento: no debe contar igual la opinión de un usuario experto (un usuario que ha aportado conocimiento que ha cristalizado) que la de un usuario ocasional (que se limita a observar el conocimiento de los demás). Tenemos en cuenta este aspecto en el sistema mediante lo que denominamos "comunidades virtuales de expertos".

Una comunidad virtual de expertos es un grupo de usuarios que son considerados expertos en uno o más temas relacionados. Las comunidades virtuales de expertos se forman a partir del árbol de conocimiento. Para cada nodo o tema, su comunidad de expertos está compuesta por los autores de documentos que han cristalizado en ese tema, en el tema del que descienden (el tema antecesor), en los temas que tienen como descendientes (sus temas sucesores o subtemas

directos) y en los temas que están a la misma altura que el primero (temas en el mismo nivel de profundidad).

Cuando iniciamos un área de conocimiento probablemente no tengamos suficiente masa crítica de participantes para el correcto funcionamiento de los procesos de cristalización de conocimiento. Esta masa crítica posiblemente no estará asegurada tampoco a largo plazo, cuando el conocimiento sea muy estable (es decir, esté muy cristalizado). Se propone para solucionar este problema un proceso de maduración por el que pasan los nodos del sistema.

Inicialmente, cuando un nodo o tema es creado (en particular cuando el nodo principal o raíz del árbol de conocimiento es creado) no hay suficientes "expertos" para formar una comunidad virtual, por lo que es necesario la figura de un "grupo de coordinación", y por tanto se trabaja en esta fase en un modo supervisado. En esta fase este grupo de coordinación se encarga de tomar decisiones que en la siguiente fase se hace de forma distribuida.

El grupo con los miembros del grupo de coordinación se forma en el momento de crear el área de conocimiento, y se pueden añadir nuevos miembros al grupo por consenso de los actuales. Una de las principales tareas de este grupo es la elaboración de la estructura de conocimiento, haciendo propuestas y opinando sobre ellas a través de un sistema de votaciones.

En esta fase se permite añadir documentos tanto a los miembros del grupo de coordinación como a los demás usuarios del sistema, los llamados usuarios colaboradores. Sin embargo, en principio sólo los miembros del grupo de coordinación tendrán la capacidad de opinar sobre los documentos a través de votaciones. Es decir, que serán estos usuarios los que puedan influir en la cristalización de documentos en esta fase. En esta fase el grupo de coordinación ejerce tareas de "consejo editorial".

El grupo de coordinación puede decidir avanzar al modo "Activo", causando su disolución y activando el mecanismo de cristalización del conocimiento mediante comunidades virtuales. En este caso cuando la aportación en forma de documento de un miembro de la comunidad cristaliza, éste recibe un cierto número de votos que podrá emplear en apoyar a otras aportaciones que estén ubicadas en la comunidad virtual donde se encuentra su documento cristalizado. Este mecanismo se inspira en la práctica habitual del "peer review" que emplean la mayor parte de las revistas científicas.

En lo referente a la cristalización de la estructura del árbol de conocimiento, en la fase activa los miembros de una comunidad virtual de un tema dado pueden proponer añadir un nuevo subtema, borrar un subtema, o mover un tema de un punto del árbol a otro. Para aprobar el cambio, se emplea de nuevo un mecanismo de votación por parte de los demás miembros de la comunidad, es decir, de los expertos que tienen contribuciones cristalizadas en el entorno del nodo.

Finalmente, la comunidad activa puede pasar a la fase "Estable" cuando apenas hay cambios en la estructura de conocimiento, pocos documentos son publicados en el sistema, y la mayoría de la actividad que hay en el nodo es de consultas. Se entiende que el conocimiento existente está muy cristalizado, y apenas hay cambios en contenidos y estructura. Sin embargo, si la actividad vuelve a aumentar se pasa a la fase anterior.

EXPERIENCIAS REALIZADAS CON KNOWCAT

Durante los últimos años se ha utilizado el sistema con diferentes comunidades de estudiantes, en varias áreas de conocimiento en la Universidad Autónoma de Madrid. Los nodos KnowCat que han tratado las áreas de conocimiento más representativas son: "Sistemas Operativos", que ha involucrado a estudiantes matriculados en la asignatura con el mismo nombre de segundo de Ingeniería Informática en los últimos cuatro años (el número de alumnos matriculados cada año es de aproximadamente 200); "Razonamiento bajo Incertidumbre", llevada a cabo durante los últimos cuatro años, por estudiantes de doctorado de Ingeniería Informática matriculados en el curso con el mismo nombre, en el que cada año suelen matricularse de 15 a 20 alumnos; y

"Matemáticas para la educación Infantil", área generada por estudiantes de Magisterio matriculados en la asignatura del mismo nombre, durante los últimos dos años (cada año participaron del orden de 130 alumnos).

El objetivo común en todo nodo KnowCat ha sido la generación de contenidos de calidad en los distintos temas del árbol de conocimiento. Las experiencias realizadas dan cierto soporte a la hipótesis de que cuando hay un número suficiente de documentos y votos a éstos por parte de usuarios "expertos", el resultado es una descripción aceptable del tema. Además en todas las experiencias realizadas también se pretendía comprobar la evolución del grado de aceptación de los documentos con el fin de determinar la correcta clasificación de documentos en cada tema.

En las dos primeras experiencias los alumnos trabajaron y profundizaron en la estructura del conocimiento, incluso más específicamente en la segunda (Razonamiento bajo Incertidumbre), donde el principal objetivo era precisamente probar el mecanismo de cristalización de la estructura jerárquica de conocimiento.

Los resultados obtenidos tras las experiencias llevadas a cabo con el sistema soportan parcialmente las hipótesis sobre las que se sustenta el diseño de KnowCat:

1. *Si un conjunto de personas expertas interactúan de manera razonable con el sistema, el resultado converge a cierto consenso.* En la mayoría de los casos los documentos en la parte más alta del ranking o baremo recibieron la mayoría de los votos (entorno al 50 % de los votos de los alumnos), es decir, hay coincidencia en el juicio de la calidad al menos para las contribuciones más destacadas. En algunos casos los documentos que describen un mismo tema tenían un grado de aceptación similar, pero eran diferenciables dos grupos de documentos, uno en la parte alta y otro en la parte baja del ranking, es decir, en estos casos al menos había consenso entre los alumnos para determinar cuáles son los mejores y cuáles los peores documentos que describen un tema.
2. *El consenso está relacionado con una medida objetiva de la calidad de las contribuciones.* Contrastando los grados de aceptación de cada documento y las posiciones que ocupaban éstos en el ranking tras las opiniones de los alumnos, con la valoración "objetiva" del profesor, casi en la totalidad de los casos, al menos dos de los tres documentos elegidos por el profesor como los mejores, estaban entre los tres primeros documentos elegidos por los alumnos.
3. *El conocimiento en el sistema está en evolución.* La mayoría de los temas de las experiencias realizadas tienen en sus primeras posiciones del ranking documentos que han sido aportados en distintos años, lo que nos confirma que es posible que documentos aportados "a posteriori" cristalicen y desbanquen a documentos que llevan más tiempo en el sistema y estaban ocupando las primeras posiciones en el ranking.
4. *La clasificación del conocimiento en forma de árbol jerárquico de temas se ha mostrado como una aproximación razonable para manejar y gestionar el conocimiento.* Debido principalmente a la naturaleza del conocimiento tratado, la estructura de conocimiento propuesta ha cumplido satisfactoriamente, en todas las experiencias, su labor de permitir clasificar y organizar el conocimiento aportado por los usuarios.

TRABAJOS FUTUROS

Quedan otros ámbitos en los que se pretende trabajar de forma inmediata, que abarcan aspectos relacionados con los puntos siguientes y se desarrollan brevemente después:

- Los modos de acceso al conocimiento y de interacción del usuario con el sistema
- El establecimiento de interacción entre instancias (nodos) del sistema distribuidas por la red y sus comunidades de usuarios
- Los procedimientos para manejar las relaciones entre documentos, instancias del sistema y comunidades, derivados de la interacción anterior

- La incorporación de instrumentos automáticos para la obtención de información sobre los documentos, los usuarios y la interacción entre ambos
- La incorporación de nuevos algoritmos de cristalización que utilicen la información obtenida con los anteriores procedimientos y la interacción con otros nodos
- La extensión de los campos de aplicación del sistema fuera del entorno académico
- El rendimiento del sistema, la administración y la seguridad
- La profundización en el estudio del desarrollo temporal de los procesos de cristalización
- El análisis de la influencia del tipo de conocimiento en la dinámica del sistema y los procesos de cristalización, más allá del conocimiento enciclopédico considerado inicialmente

Actualmente el sistema permite la existencia de múltiples nodos KnowCat, funcionando en una misma máquina o en equipos distribuidos por la red, pero no soporta ningún procedimiento de interacción entre ellos. Esta posibilidad propiciaría el establecimiento de redes de nodos distribuidos, que podrían integrar de forma automática sus conocimientos y comunidades. Para ello sería necesario el desarrollo de algoritmos de cristalización para múltiples nodos.

En este momento el sistema permite la convivencia de varias estructuras para organizar el conocimiento. Se pretende desarrollar la herramienta en el sentido que esta característica apunta, dotando al sistema de la capacidad de permitir navegar por la información y acceder al conocimiento de múltiples formas, y de presentar estos elementos de distintas maneras.

Esta capacidad se complementará con el empleo de documentos virtuales que permitirán la combinación de información incorporada por el sistema, con anotaciones, referencias y los propios documentos originales, y también la ubicación de todos los elementos en distintos puntos de las estructuras de organización del conocimiento.

Todo lo anterior se pretende conseguir con la incorporación de técnicas de la llamada web semántica, que proporcionarán un soporte adecuado para facilitar la localización, compartición e integración de los elementos del sistema, mediante la incorporación de conocimiento semántico explícito sobre ellos.

En lo referente a las unidades de conocimiento, actualmente se está considerando que las distintas versiones de cada documento son obra de los autores del original, aunque las anotaciones que eventualmente podrían motivar la generación de nuevas versiones sean aportaciones de otros miembros de la comunidad. Esto podría ser de otras maneras, como versiones comunitarias o de otros autores a partir del original de uno. Pero siempre es necesaria la adecuada gestión de la propiedad intelectual y de los derechos de autor derivados.

Por otro lado, se pretende facilitar la referencia entre documentos tomando en cuenta la indicación a partes de ellos, aunque se mantenga la consideración del artículo como unidad atómica e indivisible, mediante la estructuración interna de los documentos del sistema. De esta manera será posible referirse desde un documento a una sección de otro sin extraerla de su contexto y preservando así su significado.

Para ello se aplicarán técnicas de explotación automática de la información, que permitirán dotar a los documentos de estructura interna, facilitando la labor del autor (como la generación de resúmenes, índices o referencias), además de otros procedimientos que requieren la atención del usuario, que son las normas establecidas para regular el modo de uso del sistema (como los formatos indicados para documentos y anotaciones), y los procedimientos interactivos incorporados en el mismo (como editores y formularios para la inserción de meta-datos).

También se extenderá el conjunto considerado de interacciones implícitas de los usuarios, mediante el empleo de nuevas técnicas para la supervisión del comportamiento frente al sistema de cada uno de los usuarios en particular (documentos más accedidos, repetición y frecuencia de

las consultas, duración de las mismas...) y no de todos como individuos anónimos, tal como se hace en la actualidad.

Con el análisis de los datos obtenidos con las técnicas de los párrafos anteriores se refinarán los algoritmos de cristalización actuales para incorporar los nuevos elementos de juicio mencionados, como la interacción de cada uno de los usuarios con el sistema, las referencias entre artículos, la calidad de los usuarios que interactúan de forma implícita, etc.

Por otro lado, se proyecta ampliar los campos de aplicación de la herramienta más allá del ámbito académico (grupos de interés como ONGs o foros de debate, documentación, gestión del patrimonio...) y del tipo de conocimiento, enciclopédico, considerado inicialmente (material de referencia como problemas resueltos o ejemplos de código informático, noticias...).

También se pretende analizar los aspectos operativos del sistema, para establecer sus requerimientos y límites de capacidad en función de volúmenes de uso, refinar las herramientas de administración, monitorización y seguimiento, y desarrollar la seguridad del sistema.

Por último se tiene previsto profundizar en el estudio de la proyección temporal del proceso de cristalización del conocimiento, etapas y ciclos, considerando los dominios de aplicación de la herramienta.

Muchos de estos objetivos caben en el alcance del proyecto de investigación científica y desarrollo tecnológico "ARCADIA: Organización automática del conocimiento, análisis de datos y generación dinámica de documentos en la web semántica" (TIC2002-01948) financiado por CICYT y concedido al Departamento de Ingeniería Informática de la Escuela Politécnica Superior de la UAM, recientemente puesto en marcha.

MÁS INFORMACIÓN EN:

Página con la descripción detallada del sistema: <http://knowcat.ii.uam.es/tool/esp/index.htm>

Cobos, R, Alamán, X. **Creating e-books in a distributed and collaborative way.** *Journal of Electronic Library on Electronic book for Education*. Vol. 20, no. 4, May 2002: pp. 288-295.

Cobos, R, Esquivel, J., Alamán, X. **Herramientas informáticas para la Gestión del Conocimiento: un estudio de la situación actual.** *Revista de la Asociación de Técnicos de Informática, Novatica*, special issue on Knowledge Management and Information, nº 155, January-February, 2002: pp. 20-26.

Cobos, R., Alamán, X. **KnowCat: a Knowledge Crystallisation Tool.** Proceedings of the International Conference e-Business and e-Work (e-2000). Madrid, October 17-19, 2000. In: *E-business: Key Issues, Applications and Technologies*, Stanford-Smith, B. y Kidd, P. T. (eds). IOS Press: pp. 374-380.

Alamán, X. y Cobos, R. **KnowCat: a Web Application for Knowledge Organization.** Proceedings of the International Workshop on World-Wide Web and Conceptual Modeling (WWWCM'99), in conjunction with the 18th International Conference on Conceptual Modeling (ER'99). Paris, November 15-18, 1999. In: *Lecture Notes in Computer Science 1727*, P.P. Chen et.al. (eds). Springer: pp. 348-359.