



El Proyecto Europeo Dicode: una Infraestructura *Cloud* para potenciar la colaboración y toma de decisiones en entornos de computación intensiva

Los entornos colaborativos y de toma de decisiones están frecuentemente asociados con grandes cantidades de datos de todo tipo, provenientes de diversas fuentes, cuya baja relación señal-ruido hace difícil abordar el problema. Además, estos datos pueden variar en subjetividad e importancia, incluyendo desde opiniones individuales y estimaciones hasta prácticas ampliamente aceptadas, mediciones indiscutibles o resultados científicos. Sus tipos pueden presentar distintos niveles en lo que se refiere a la comprensión humana y la interpretación que hacen las máquinas.

En la actualidad, grandes volúmenes de datos pueden ser añadidos sin esfuerzo a una base de datos; los problemas comienzan cuando queremos estudiar y explotar los datos acumulados durante unas pocas semanas o meses, y analizarlos provechosamente para tomar una decisión. En situaciones más complejas es necesario identificar, comprender y aprovechar los patrones existentes en los datos; es necesario integrar grandes volúmenes de datos provenientes de múltiples fuentes y extraer de ellos nuevo conocimiento que nunca surgiría de una simple inspección manual o del análisis de una única fuente de datos.

Partiendo de estos problemas, el nuevo proyecto europeo Dicode (<http://dicode-project.eu/>) pretende facilitar y aumentar la colaboración y toma de decisiones en entornos cognitivamente complejos y de computación intensiva. Haciendo uso de avanzados métodos y tecnologías de computación de alto rendimiento y procesamiento de grandes volúmenes de datos—p.e. *Cloud computing*, *MapReduce*, *Apache Hadoop* o *Column databases* permitirá buscar, agregar y analizar datos existentes en diversas, extremadamente grandes y continuamente cambiantes fuentes de datos. Los servicios desarrollados e integrados en el contexto del proyecto Dicode serán publicados bajo licencia *Open Source*.

El proyecto Dicode parte en su estrategia de varias situaciones:

- La creciente presencia del *Cloud computing*, tanto en la industria como en los entornos académicos, como una plataforma escalable de distribución y consumo de servicios. Dichos servicios proporcionan utilidades cada vez más potentes. Comparado con sus predecesores—p.e. *Grid computing*, *Utility computing*—el *Cloud computing* está mejor posicionado en términos de viabilidad económica, rentabilidad, escalabilidad, fiabilidad, interoperabilidad e implementaciones *Open Source*
- Existen numerosos avances en el desarrollo de tecnologías y plataformas escalables para la minería de datos como *MapReduce*, *Apache Hadoop* y *Apache Mahout*. La mayor parte de ellos aprovechan el paradigma del *Cloud computing*. Asimismo, las tecnologías de minería de textos—p.e. el reconocimiento y desambiguación de nombres de entidades, extracción de relaciones o la minería de opinión—han alcanzado un nivel tal que, por primera vez es viable, en la práctica, aplicar tecnologías semánticas a colecciones muy grandes de datos, permitiendo capturar una cantidad de conocimiento sin precedentes a partir de textos no estructurados
- En paralelo, existen grandes avances en el desarrollo de aplicaciones colaborativas y de soporte a la toma de decisiones, principalmente aprovechando las características y tecnologías de la Web 2.0
- Aunque resultan útiles para problemas y entornos concretos, los avances anteriormente citados presentan una serie de limitaciones e ineficiencias cuando son aplicados en

configuraciones de computación intensiva y cognitivamente complejas en entornos colaborativos y de soporte a la toma de decisiones

Basada en avances actuales, la solución prevista en el proyecto Dicode combinará las capacidades de razonamiento humanas y de las máquinas. Puede verse como un innovador entorno de trabajo que incorpora y orquesta un conjunto de servicios interoperables (Figura 1) que reduce la complejidad y sobrecarga de datos en los puntos de decisión críticos a un nivel manejable, permitiendo a las partes interesadas ser más productivas y concentrarse en actividades más creativas e innovadoras.

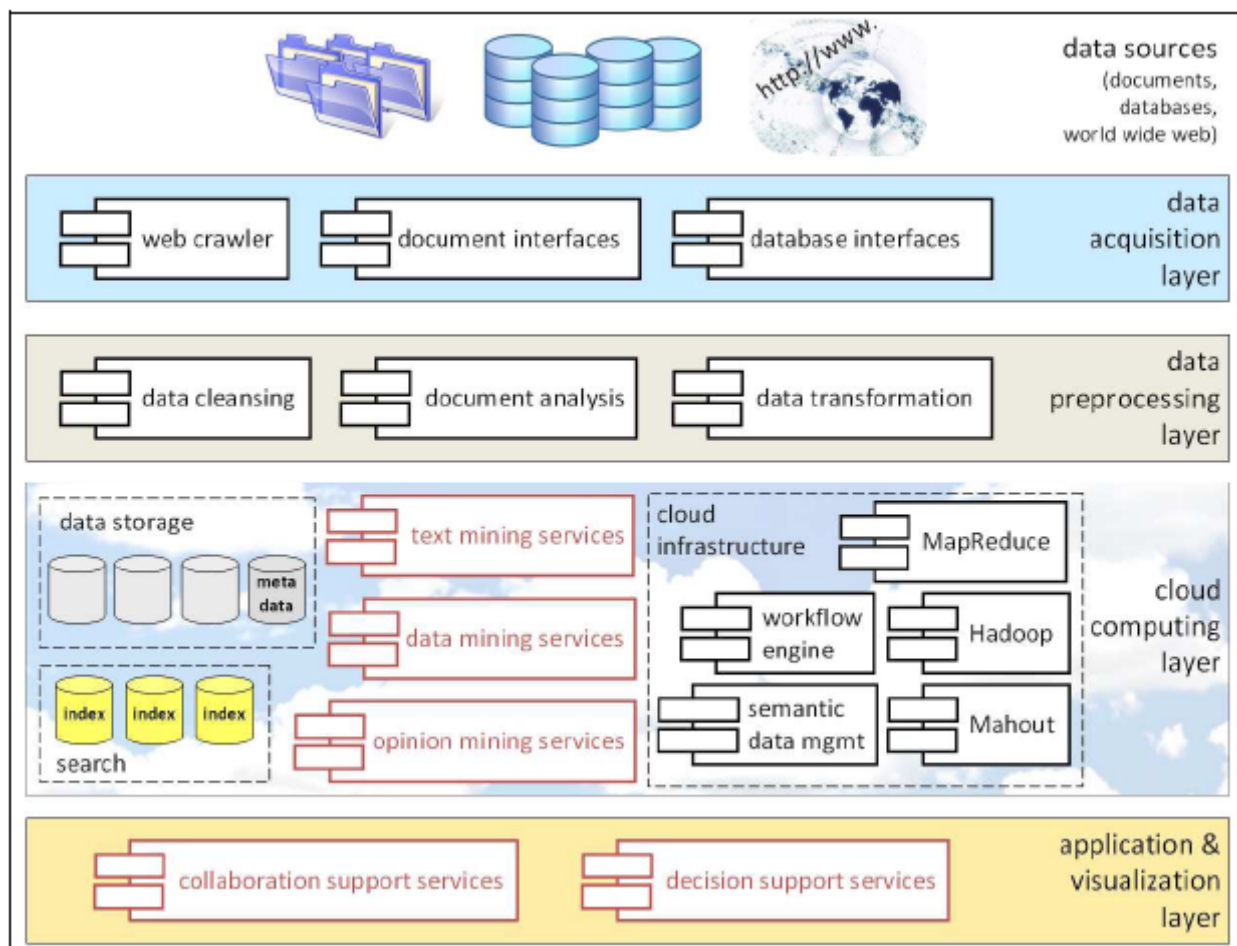


Figura 1: La Arquitectura Dicode y Conjunto de Servicios

El éxito de los objetivos del proyecto Dicode será validado a través de tres casos de uso. Estos casos fueron elegidos para demostrar la transferibilidad de las soluciones Dicode en diferentes escenarios colaborativos y de toma de decisiones, asociados con diversos tipos y fuentes de datos, cubriendo así toda la gama de características y funcionalidades previstas en dichas soluciones.

Los casos de uso son:

1. *Descubrimiento de conocimiento Clínico-Genómico para la Investigación.* Este caso de uso demostrará cómo la arquitectura Dicode puede dar soporte a las investigaciones científicas clinico-genómicas en la actual era post-genómica. La necesidad de explorar, evaluar, diseminar y difundir colaborativamente los hallazgos y resultados científicos más relevantes es crucial actualmente. Para lograr dicho objetivo, Dicode desarrollará un entorno integrado para el descubrimiento de conocimiento clinico-genómico y toma de decisiones cuyo objetivo es la identificación y validación de modelos predictivos clinico-genómicos y biomarcadores. El caso de uso se basa en la perfecta integración tanto de fuentes de datos clinico-genómicas heterogéneas como en técnicas analíticas avanzadas desarrolladas por Dicode.

2. *Proyecto piloto para el estudio informático del tratamiento de la Artritis Reumatoide*. Este caso de uso aprovechará los servicios de Dicode para distribuir información relevante a las comunidades de médicos y pacientes en el dominio de la Artritis Reumatoide (AR). Los ensayos de tratamientos de AR se realizarán en centros académicos de investigación en colaboración con la industria farmacéutica. Cada ensayo evaluará la eficacia del tratamiento para la AR analizando la afección en las muñecas (y posiblemente otras articulaciones). Los servicios de Dicode se utilizarán para desarrollar una nueva forma de trabajar efectiva y colaborativa orientada hacia la toma de decisiones, involucrando a diferentes personas (radiólogos, médicos, etc.).
3. *Minería de Opiniones a partir de datos no estructurados de la Web 2.0*. Actualmente es fundamental para las empresas conocer lo que se dice sobre sus servicios o productos. Con las herramientas actuales encontrar quién y qué se está diciendo es, literalmente, como buscar una aguja en un pajar de información no estructurada. Mediante este caso de uso, el objetivo es validar los servicios de Dicode para el análisis automático de voluminosas cantidades de información no estructurada. Los datos se obtendrán principalmente de la Web (*blogs*, foros y noticias) y también haciendo uso de diferentes APIs de varias plataformas de la Web 2.0, tales como plataformas de *micro-blogging* (*Twitter*) y redes sociales (*Facebook*)

El proyecto Dicode está financiado por la Unión Europea dentro del 7º Programa Marco (Coste total: 3.510.000 €, financiación: 2.600.000 €). El proyecto comenzó el 1 de septiembre de 2010 y tiene una duración de 36 meses. Los socios del consorcio Dicode son: Research Academic Computer Technology Institute (project coordinator), University of Leeds, Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., Universidad Politécnica de Madrid, neofonie GmbH, Image Analysis Ltd, Biomedical Research Foundation - Academy of Athens, and Publicis Frankfurt GmbH.

Para contacto en español: UPM (Guillermo de la Calle gcalle@infomed.dia.fi.upm.es y Víctor Maojo vmaojo@fi.upm.es)

Dicode en Twitter: http://twitter.com/DICODE_EU

Dicode en Facebook: <http://www.facebook.com/people/Dicode-Eu/100001390513581>

Personas de contacto:

Nikos Karacapilidis

Research Academic Computer Technology Institute
26504 Rio Patras, Greece
Tel: +30 2610 960305
E-mail: karacap@cti.gr

Stefan Rüping

Fraunhofer Institute for Intelligent Analysis and Information Systems
53757 Sankt Augustin, Germany
Tel: +49 2241 143512
E-mail: stefan.rueping@iais.fraunhofer.de

Scott Robinson

neofonie GmbH
Robert-Koch-Platz 4, 10115 Berlin, Germany
Tel: +49 30 246 27 562
E-mail: scott@neofonie.de