

Hacia una plataforma de interoperabilidad para aplicaciones biológicas



Guzmán Llambías

Abril 2010

Agenda

- ¿Qué son las aplicaciones biológicas?
- ¿Qué problemas existen?
- Una propuesta de cómo resolverlos



¿Qué son las app. biológicas?

- ❑ Es todo aquel software utilizado por los científicos para hacer sus experimentos
 - Bases de datos y servicios de acceso
 - GenBank, EBI, DDBJ
 - Herramientas analíticas
 - Blast, ClustalW, scripts en lenguaje R
 - Visualizadores
 - ...

- ❑ Los científicos hacen sus experimentos utilizando este tipo de aplicaciones.
 - Realizan experimentos in-silico!



Experimentos in-silico

- ❑ Es un procedimiento que involucra recursos computacionales locales o remotos para el desarrollo de pruebas que permitan verificar ciertas hipótesis, derivar resultados, buscar patrones o demostrar hechos conocidos

- ❑ En bioinformática, estos recursos pueden ser:
 - Repositorios de información
 - GenBank (USA), EBI (Europa), DDBJ (Japón)
 - Herramientas computacionales
 - BLAST, ClustalW
 - ...



Ejemplo

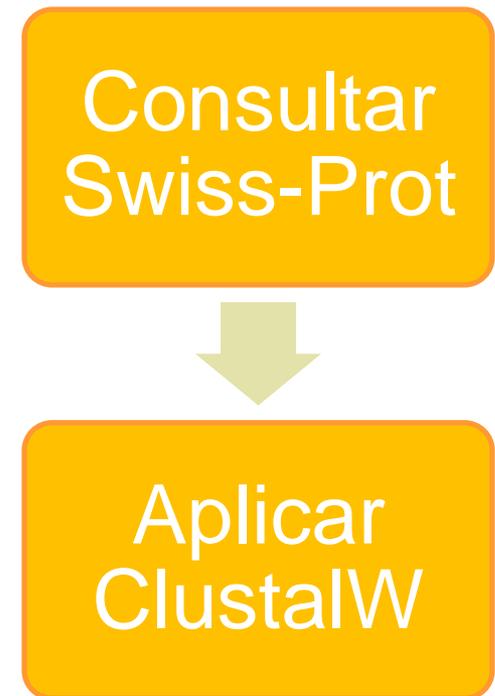
- Objetivo:
 - Investigar la evolución de las relaciones entre proteínas

- Experimento:
 1. Adquirir una secuencia de aminoácidos del repositorio Swiss-Prot
 2. Aplicar el algoritmo de ClustalW para alinear e identificar patrones entre las secuencias



Ejemplo

- ❑ Objetivo:
 - Investigar la evolución de las relaciones entre proteínas
- ❑ Experimento:
 1. Adquirir una secuencia de aminoácidos del repositorio Swiss-Prot
 2. Aplicar el algoritmo de ClustalW para alinear e identificar patrones entre las secuencias
- ❑ Los experimentos in-silico, siguen una estructura de workflow
 - Workflows Científico!



Experimentos in-silico

- ❑ Primera generación
 - Cut & paste entre aplicaciones Web
 - Catalogada como “Tortura medieval”

- ❑ Segunda generación
 - Scripts en lenguajes de programación
 - Consultas Web y técnicas Screen Scraping
 - Open bioinformatics: BioPerl, BioJava, BioPython...

- ❑ Tercera generación
 - Herramientas basadas en middleware de alto nivel
 - Web Services + Workflows + ...



- Un proyecto de middleware en e-Science

- Principal objetivo:
 - desarrollo y provisión de herramientas y aplicaciones basadas middleware de alto nivel que les permita a los científicos realizar sus experimentos abstrayéndose de conceptos computacionales avanzados.

- Herramientas desarrolladas
 - Taverna: diseño, edición y ejecución de workflows
 - myExperiment: una plataforma orientada a la Web 2.0 para compartir workflows y datos
 - BioCatalogue: permite anotar y catalogar servicios



Proyecto myGrid

Diseñar, editar
y ejecutar
workflows



Buscar y
catalogar
servicios

BioCatalogue ^{beta} 
"The Life Science Web Service Registry"

The "my experiment" logo consists of the word "my" in white text on a red rectangular background, followed by the word "experiment" in green text on a white rectangular background.

Buscar y
publicar
resultados

myGrid 



- ❑ Taverna es una potente herramienta desarrollada por el proyecto *myGrid*, para el modelado, edición y ejecución de workflows

- ❑ Permite a los científicos realizar sus experimentos abstrayéndose de conceptos computacionales avanzados:
 - Comunicación con Web Services
 - Descubrimiento de recursos (herramientas y BDs)
 - Registro de *provenance* (linaje de los experimentos)
 - Formateo de datos
 - ...



Sin embargo...

- ❑ Taverna no resuelve adecuadamente todas las tareas de interés para los científicos, algunas de las cuales siguen siendo resueltas en forma manual o fuera de los workflow
 - Transformación de formatos de datos
 - Interoperabilidad con otras herramientas de workflow
 - Adaptación en caso de errores o excepciones

- ❑ Algunos de estos problemas ya han sido abordados con éxito en las plataformas empresariales (SOA, ESB, etc)



- ❑ Incorporar una nueva capa de abstracción entre los workflows y servicios que se encargue de resolver los problemas de infraestructura

- ❑ Es de interés explorar el aporte que podrían hacer características típicas de las plataformas empresariales
 - Asincronismo
 - Comunicaciones orientadas a eventos
 - Ruteo inteligente de mensajes
 - ...



Enterprise Service Bus (ESB)

- ❑ Es un tipo de middleware que provee la infraestructura necesaria para el desarrollo de aplicaciones siguiendo el modelo SOA
- ❑ Principales características
 - Mensajería
 - Mecanismos de mediación
 - Direccionamiento inteligente de mensajes
 - Orientación a eventos
- ❑ Lema del ESB
 - Separar la lógica de integración y comunicación de la lógica de negocio.



Bosquejo de la solución



Nueva generación de workflows

- Basado en plataformas de middleware avanzado de gran abstracción compuesto por:
 - Herramientas de workflows
 - Repositorios de información
 - Herramientas analíticas
 - ESB
 - Metadatos
 - Web 2.0
 - ...

- Los científicos se concentran en sus **datos**, sus **procesamientos** y sus **resultados**. No en tareas de infraestructura e interoperabilidad



En qué estamos...

- ❑ Maestría en curso
- ❑ Proyecto grado
 - Integración de Taverna a JBoss ESB
 - Aplicación de mecanismos de mediación interesantes
- ❑ Buscando interesados con quienes colaborar

- ❑ Antecedentes y contexto
 - Tomamos como base la experiencia de actividades con el Instituto Pasteur
 - LIMS, workflows, calidad de datos
 - Cursos y actividades en la Maestría de Bioinformática



Trabajo futuro

- ❑ Estudiar con mayor profundidad Taverna y software afines
- ❑ Analizar con mayor profundidad problemas de integración entre aplicaciones biológicas
- ❑ Analizar el impacto de mecanismos asincrónicos, manejo de eventos y ruteo inteligente de mensajes en los workflows
- ❑ Desarrollo de una plataforma de interoperabilidad para aplicaciones biológicas



Preguntas





Contacto

Guzmán Llambías
gllambi@fing.edu.uy