

CU9: VO y Minería de Datos

Balance global y futuras actuaciones

Resumen:

- [X. Luri] Contexto. CU9, status, descripción de actividades y marco general de diseño de la CU.
- [P. Osuna] Descripción de infraestructura y tecnologías consideradas para core system.
- [Discusión]
 - Qué entendemos como actividades de core system y qué entra en la caja de herramientas avanzadas.
 - Qué casos de uso van a dictar los requisitos sobre el core system y qué prioridades vamos a asignarle
- [D. Fustes] Ejemplo de herramienta de MD
 - Mapas autoorganizados.

Resumen de discusión I:

- Existen dos tipos de usuarios:
 - Usuarios típicos de una base de datos
 - Usuarios de técnicas avanzadas de análisis.
- Paradigma 80-20: podemos esperar que el 20% de los usuarios necesiten y trabajen sobre el 80% de los datos y viceversa.
- Las técnicas avanzadas implementadas en CU9 deben ser definidas de acuerdo con los casos de uso. Workshop 2013? 2014?

• Resumen de discusión II:

- Un caso de uso típico: comparación de modelos Galácticos con la base de datos de Gaia. ¿Hay una barra? ¿Dos? ¿Qué ángulo forman? ¿Existe un disco grueso? ¿Cuáles son sus propiedades? ¿SFR? ¿IMFs?
- Técnica: modelos bayesianos jerárquicos construidos con técnicas de muestreo estocástico.
- Prueba de concepto con GUMS 11 (ESAC, UB, UNED)

Resumen de discusión III:

- Hay que posibilitar a los usuarios avanzados una infraestructura de cálculo: Hadoop, MapReduce, ESAC/CESGA/CESCA/CESVIMA acuerdos?, Mahout
- ¿Comités de asignación de tiempo?
- Enlace con BBDD externas. Por ejemplo, seleccionar subconjunto de datos de acuerdo a criterios que incluyan datos de Gaia, 2MASS, Gaia-ESO y aplicar técnicas de MD al resultado.

Resumen de discusión IV:

- Una reunión de arranque para CU9 hacia mitad de 2013.
- Sería necesaria una reunión previa de los grupos españoles con participación en CU9 para coordinar los esfuerzos.