

# TRABAJOS ASIGNATURA SAR

Curso 2020-21

**OBJETIVO:** preparar y realizar una presentación pública sobre un tema relacionado con la asignatura.

**PROCEDIMIENTO:**

1. Los alumnos formarán grupos de 4 componentes.
2. Cada grupo nos comunicará a través del **correo interno de la asignatura** dirigido a **Rol Profesor**, con fecha tope el **12 de marzo**:
  - a. Los componentes del grupo (nombre completo de todos los componentes)
  - b. El nombre del representante del grupo y su correo electrónico
  - c. La lista ordenada por orden de preferencia de las propuestas de trabajos publicadas en este documento (por ejemplo: T5, T6, T1, ...)
3. El **22 de marzo** publicaremos la asignación definitiva de trabajos a grupos.
4. El representante de cada grupo deberá subir hasta el día **14 de junio** al PoliformaT en una tarea que se habilitará para ello:
  - a. La presentación .ppt o/y .pdf del trabajo elegido. A continuación, se muestra un esquema de una presentación genérica. La presentación deberá contener los nombres de los componentes del grupo.
  - b. Un vídeo de unos 15 minutos en total de la presentación realizada por todos los componentes del grupo, cada componente debe participar en una parte de la presentación. Si se presentan problemas en la subida debidos al tamaño de los ficheros, siempre se puede subir un enlace a una carpeta externa con los vídeos.
5. En la lista de propuestas de trabajos que se relacionan a continuación, se indica el título de los capítulos de libro o artículos que deben servir de referencia principal para la preparación del trabajo. Se puede ampliar con otros documentos. En función del interés del grupo o de la extensión del capítulo de referencia, se puede trabajar sobre una parte de los contenidos del capítulo dejando otros aspectos sin tratar.

## **ESQUEMA GENERAL DE LA PRESENTACIÓN**

- Título y autores
- Índice
- Introducción (se introduce el problema que se va a abordar)
- Desarrollo (se organiza en puntos y se desarrollan)
- Resumen de puntos más interesantes
- Bibliografía

## **CRITERIOS DE EVALUACIÓN DE LA PRESENTACIÓN**

### **ORGANIZACIÓN DEL DISCURSO: Introducción**

Se ha realizado una buena introducción del tema

Se ha presentado un índice de los contenidos a desarrollar

La introducción ha servido para ganar la atención de la audiencia y despertar el interés por el tema

### **ORGANIZACIÓN DEL DISCURSO: Desarrollo**

La secuencia de contenidos ha sido lógica y coherente

Se han subrayado los conceptos a destacar

Se ha mostrado una buena capacidad de síntesis

### **ORGANIZACIÓN DEL DISCURSO: Conclusiones**

Al final de la exposición, se ha incluido un resumen de la misma (conclusiones)

Se ha presentado una bibliografía adecuada al tema

## **VALORACIÓN DE LA EXPOSICIÓN**

Se ha ajustado al tiempo establecido

Se ha transmitido firmeza y seguridad en el discurso

La exposición ha resultado clara e inteligible (usa un vocabulario adecuado y variado, se expresa con corrección, etc.)

La entonación y el volumen han facilitado el seguimiento del discurso

Los recursos utilizados se han empleado de manera adecuada (diseño, fuentes, colores de la presentación, etc)

## PROPUESTAS DE TRABAJOS:

**T1:** Language Models for Information Retrieval (Cap 12, “An Introduction to Information Retrieval”, C. Manning).

Se trata de desarrollar una labor de síntesis más alguna aportación nueva complementando alguno de los aspectos del uso de técnicas de modelado del lenguaje para la recuperación de información.

**T2:** Multimedia Information Retrieval (Cap 14, “Modern Information Retrieval”, R. Baeza).

Se trata de desarrollar una labor de síntesis del capítulo, dejando sin tratar el apartado final de compresión. Se puede realizar alguna aportación nueva relativa por ejemplo a tendencias actuales en recuperación de información multimedia.

**T3:** Recuperación de Imagen (Cap 17, “Recuperación de Información: un enfoque práctico y multidisciplinar”, F. Cacheda, J.M. Fernández y J.F. Huete).

Se trata de desarrollar una labor de síntesis del capítulo centrando el trabajo en algunos de los tipos de representación del contenido.

**T4:** Relevance feedback and query expansion (Cap. 9, “An Introduction to Information Retrieval”, C. Manning).

Se trata de desarrollar una labor de síntesis más alguna aportación nueva, complementando por ejemplo alguna de las estrategias de retroalimentación de la relevancia o bien sobre algún método de reformulación de las consultas que se exponen en este capítulo.

**T5:** XML Retrieval (Cap. 10, “An Introduction to Information Retrieval”, C. Manning).

Se trata de desarrollar una labor de síntesis más alguna aportación nueva, complementando por ejemplo alguno de los aspectos como el modelo de recuperación XML, evaluación de la recuperación de XML o cualquier otro que resulte de su interés.

**T6:** Detección de Plagio en texto (“Plagiarism Detection Using Stopword n-grams”. Stamatos E. Journal of the American Society for Information Science and Technology, December, 2011, <http://onlinelibrary.wiley.com/doi/10.1002/asi.21630/pdf> “Detección automática de plagio: de la copia exacta a la paráfrasis”. Barrón A., Vila M., Rosso P. Panorama actual de la lingüística forense en el ámbito legal y policial: Teoría y práctica. 2010, [http://users.dsic.upv.es/~proso/resources/BarronEtAl\\_JLF10.pdf](http://users.dsic.upv.es/~proso/resources/BarronEtAl_JLF10.pdf)

Se trata de desarrollar una labor de comprensión y síntesis de los trabajos referenciados sin entrar en los detalles de la paráfrasis.

**T7:** Recommendation Systems (Cap 15, “Recuperación de Información: un enfoque práctico y multidisciplinar”, F. Cacheda, J.M. Fernández y J.F. Huete) y (Cap 9, “Mining of Massive Datasets”, A. Rajaraman).

Se trata de desarrollar una labor de síntesis más una aportación nueva, complementando por ejemplo algunos aspectos sobre la obtención de datos para la matriz de utilidad.

**T8:** Question Answering (Cap 23, “Speech and Language Processing”, Jurafsky and Martin (draft chapters in progress 2020) <http://web.stanford.edu/~jurafsky/slp3/>).

Se trata de hacer un trabajo del tema donde se valorará la capacidad de síntesis y comprensión del mismo.

**T9:** Summarization.

Se trata de desarrollar una labor de síntesis sobre los métodos de resumen automático de documentos. Para ello se usará el artículo “Text summarisation in\_progress. A literatura review” Elena Lloret, Manuel Palomar (Artificial Intelligence Review V.37, pp-1-41, 2012), que proporciona una visión general de los métodos existentes, junto con el artículo “Text Summarization of Turkish Texts using Latent Semantic Analysis” Makbule Gulcin Ozsoy, Ilyas Cicekli, Ferda Nur Alpaslan, Proc. of 23rd Coling 2010, pages 869–876, que presenta un forma concreta de realizarlo.

**T10:** Twitter Sentiment Analysis. [The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation](#)

Se trata de desarrollar una labor de síntesis sobre el artículo propuesto. Se valorará la capacidad de síntesis y comprensión del texto.

**T11:** Recuperación de Información entre lenguas distintas (Exploiting Wikipedia for cross-lingual and multilingual information retrieval, Philipp Sorg, Philipp Cimiano, Data & Knowledge Engineering, Volume 74, April 2012 // “Cross-lingual Information Retrieval with Explicit Semantic Analysis” Philipp Sorg, Philipp Cimiano, Working Notes for the CLEF 2008)

<http://www.sciencedirect.com/science/article/pii/S0169023X12000213>

Se trata de hacer un trabajo del tema donde se valorará la capacidad de síntesis y comprensión del mismo.

**T12:** Similitud de textos

“Una aproximación al uso de Word embeddings en una tarea de similitud de textos en español”. T López-Solaz, JA Troyano, FJ Ortega y F Enríquez, Procesamiento del Lenguaje Natural, Revista nº 57, 2016, págs. 67-74.

<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5338>).

“Similitud español-inglés a través de word embeddings”. Fernando Enríquez, Fermín Cruz, F. Javier Ortega, José A. Troyano, Procesamiento del Lenguaje Natural, Revista nº 59, septiembre de 2017, pp. 31-38.

<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5490>

Se trata de hacer un trabajo del tema donde se valorará la capacidad de síntesis y comprensión del mismo.

**T13:** Dbpedia (“DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia”. Lehmann et al., Semantic Web 1 (2012) 1–5).

[http://svn.aksw.org/papers/2013/SWJ\\_DBpedia/public.pdf](http://svn.aksw.org/papers/2013/SWJ_DBpedia/public.pdf)

Se trata de hacer un trabajo del tema donde se valorará la capacidad de síntesis y comprensión del mismo.

**T14:** ConceptNet (ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. Robyn Speer, Joshua Chin, and Catherine Havasi. 2017, In proceedings of AAAI)

<https://arxiv.org/abs/1612.03975>

Se trata de hacer un trabajo del tema donde se valorará la capacidad de síntesis y comprensión del mismo.