

**Título del proyecto:** TEMANTEX: Análisis temporal de textos.

**Responsable:** Dina Wonsever

**Servicio:** Facultad de Ingeniería, Instituto de Computación

**Resumen:**

En el desarrollo del proyecto TEMANTEX se elaboraron esquemas validados de anotación de eventos, expresiones temporales y marcas temporales para textos en español. Se realizó la anotación manual de un corpus según todos estos esquemas. En el proceso de anotación participaron lingüistas especialmente contratados y estudiantes de Lingüística; cada texto tiene una doble anotación. Se realizó un estudio particular de las expresiones temporales que se concretó en la definición de un lenguaje de descripción de expresiones temporales (LDT). En la parte de descripción de eventos y relaciones se trabajó con un esquema fuertemente inspirado en el esquema TimeML, con algunas diferencias resultado de su adaptación al español y, como agregado fundamental, un atributo llamado *factividad* que da cuenta del estado de realizado o no de un evento. Por el contrario, la propuesta para expresiones temporales difiere radicalmente, tanto en su alcance como en el tipo de representación propuesto, de otras propuestas en el estado del arte. Se generó un corpus anotado según los esquemas definidos, con unos 55.000 tokens, libremente disponible. Se construyó una herramienta de software basada en reglas para la recuperación e interpretación de expresiones temporales. La evaluación sobre un pequeño corpus de testeo arroja, para la medida F, un 85,1% en reconocimiento y un 80% en interpretación. Se realizaron experimentos de reconocimiento automático, aplicando dos métodos de aprendizaje (*Conditional Random Fields, CRF* y *Support Vector Machines, SVM*) para el reconocimiento de eventos, con resultados semejantes a los reportados en el estado del arte en trabajos similares para el inglés (76,72% en CRF y 80,34% en SVM, para la medida F).