



ANÁLISIS DE DATOS NUMÉRICOS Y TEXTUALES PARA INTELIGENCIA COMPETITIVA DE NEGOCIOS

Director: **Russo, Alfredo**
arusso@untref.edu.ar

Comprende el desarrollo de aplicativos en de software libre (de preferencia) o propietario (planillas electrónicas) existente para la creación de conocimiento experimental en las áreas mencionadas. De acuerdo con las necesidades locales, la aplicación de estas técnicas requiere la realización previa de tareas conocidas como ETL (extracción, transformación y carga) de los datos es necesaria como etapa previa de la interpretación de los mismos. En trabajos anteriores se han puesto a punto las técnicas de análisis de datos con el uso de programas de Software como Weka y Pentaho para la exploración numérica, falta aún completar una fase de creación y adaptación de programas que procesen lenguaje XML y similares con el fin de poder convertir texto normal en texto etiquetado. La técnica experimental es similar a la que se usa en redes neurales: una parte del texto se usa para extraer los conjuntos significativos de palabras que lo caracterizan y construir el lexikon para interpretar el resto del texto. También será explorada en forma experimental la consulta a expertos para la creación de lexikon. Se utilizará la herramienta de software libre R con el reciente Rattle y Yale, con Word Vector, así como GATE, un programa de la Universidad de Sheffield, para la Minería de Textos. Rapid Miner tienen funcionalidades de ETL para la preparación previa de datos.

ANALYSIS OF NUMERICAL AND TEXTUAL DATA FOR COMPETITIVE BUSINESS INTELLIGENCE

Director: **Russo, Alfredo**
arusso@untref.edu.ar

It includes the development of existing free software (preferably) or proprietary (spreadsheets) applications for the creation of experimental knowledge creation in these areas. According to local needs, the application of these techniques requires the prior execution of tasks known as ETL (extraction, transformation and loading) of data as a prior step to their interpretation. Previous works have perfected the techniques of data analysis using software programs such as Weka and Pentaho for numeric exploration. Yet to be completed is a phase of creation and adaptation of programs that process XML and similar languages to able to convert normal text into labeled text. The experimental technique is similar to that used in neural networks: one part of the text is used to extract significant sets of words that characterize it to build lexicon to interpret the rest of the text. We will also experimentally explore consulting with experts to create lexicon. The R free software tool will be used with the recent Rattle and Yale, with Word Vector and GATE, a program from the University of Sheffield, for Text Mining. Rapid Miner has ETL capabilities for the prior preparation of data.