Desarrollo y aplicación de técnicas de extracción de información en Data Science

Objetivos y finalidad

Hoy en día, las técnicas de extracción de información encuentran algunas limitaciones que se deben, entre otras cosas, al volumen, la complejidad, la heterogeneidad, la necesidad de consistencia, velocidad de procesamiento y comunicación de los datos. Resolver esas limitaciones en problemáticas específicas, plantea nuevos desafíos, que llevan a proponer nuevos métodos y técnicas o adaptar los existentes.

El uso de datos masivos se extiende a áreas de lo más diversas, desde las relaciones humanas a la medicina, industrias de servicios y manufacturas, agricultura y hasta al ámbito jurídico. Esta nueva perspectiva hace evidente la limitación relacionada con el volumen, pero también conduce directamente a las ideas de diversidad y complejidad presentes en los diferentes tipos de datos.

El objetivo general propuesto para este proyecto es el estudio, el desarrollo y la aplicación de técnicas de extracción de información en diferentes problemas enmarcados en el área conocida como “Data Science” o Ciencia de Datos. Estas técnicas se aplican en diferentes contextos, entre ellos: el Modelado y Procesamiento de Datos Convencionales y no Convencionales, el Reuso Semántico de Modelos de Requisitos, la Clusterización en Imágenes Multidimensionales, los Sistemas Médicos de Asistencia al Diagnóstico y al Tratamiento y los Sistemas de Información Aplicados a la Cadena Productiva.

Objetivos específicos

A continuación se presentan los objetivos específicos de cada una de las áreas de aplicación mencionadas en la sección anterior:

Modelado y Procesamiento de Datos Convencionales y no Convencionales

Las fases de modelado conceptual, lógico y físico de datos son fundamentales para el desarrollo de cualquier sistema de información y en particular para el proceso de diseño de bases de datos. Para lograr un diseño completo y exacto, se necesitan metodologías que permitan representar y trasladar todos los aspectos relevantes del Universo de Discurso (UdeD) al Esquema Físico de Datos (EFD) [Badía 2011] [Codd 1990] [Elmasri & Navathe 2010] y que se realice un mapeo directo entre el mundo real percibido y su representación, sin distorsión o ambigüedad [Pieris 2013-1] [Pieris 2013-2] [Pieris 2013-3]. En la investigación interdisciplinaria surge la necesidad de modelar y procesar datos con características particulares; aparecen de este modo nuevas perspectivas que difieren ampliamente de aquellas asociadas a los datos convencionales. En este contexto, se trabajará con datos morfométricos, con datos de señales acústicas generadas por vehículos submarinos autónomos y con datos de objetos digitales producidos al transformar fuentes documentales de soportes diversos (gráficos, sonoros, audiovisuales, escritos) entre otros [Torcida et al. 2014] [Torcida et al. 2016] [Márquez et al. 2015] [Ferraggine 2016] [Villar et al. 2017]. Por lo expuesto anteriormente, se propone:

- Continuar con la definición e implementación del proceso automatizado para transformación del modelo conceptual de datos al EFD, analizando:
  - Factores que influyen en el modelado de relaciones ternarias: representación de todas las propiedades de las relaciones ternarias.
  - Expresiones formales de la sintaxis de reglas de transformación.
  - La incorporación de aspectos semánticos de big data.
- Continuar con el análisis de características de modelado de datos morfométricos, desarrollando metodologías para su procesamiento.
- Comenzar el análisis de características de datos de señales acústicas producidas por vehículos submarinos autónomos, desarrollando posteriormente metodologías para su procesamiento.
- Comenzar el análisis de características de objetos digitales producto de la transformación de fuentes documentales.

Técnicas Cuantitativas Orientadas al Reuso Semántico de Modelos de Requisitos

Los modelos Léxico Extendido del Lenguaje (LEL), Escenarios Actuales (EA), Escenarios Futuros (EF) y Especificación de Requisitos (SRS) son construidos para dar soporte al proceso de definición de los requisitos del software [Kaplan et al. 2013] [do Prado Leite et al. 2004]. Sin embargo, su mera observación objetiva muestra que contienen más información que la que explícitamente se indica. Para poder extraer ésta y muchas otras informaciones de los modelos del proceso de Ingeniería de Requisitos (IR), se aplican técnicas basadas en estrategias cuantitativas y/o cualitativas. Es así que, en este proyecto se propone:

- Continuar con el análisis del LEL, así como EA y EF, sistematizando el metaconocimiento adquirido, con el fin de su reuso semántico en etapas posteriores del desarrollo de software.
- Crear mecanismos que permitan facilitar la visualización de los agrupamientos presentes en los grafos de los LEL’s de diferentes casos de estudio, delimitando automáticamente dichos agrupamientos.
- Extender la aplicación de los mecanismos obtenidos al resto de los modelos del proceso de IR.

Clusterización en Imágenes Multidimensionales

Obtener información de imágenes satelitales requiere ordenar y agrupar una gran cantidad de datos multidimensionales. Las técnicas de Clustering implementadas en computadoras hasta la fecha brindan soluciones aceptables, sin embargo tienden a ser lentos. Esto hace que no sean adecuados para el tratamiento de grandes bases de datos o que puedan requerir intervención humana, restandoles automaticidad y autonomía. Además, algunas de estas técnicas dependen en gran medida de las inicializaciones específicas y contienen pasos no deterministicos, lo que lleva a un largo proceso de prueba y error hasta encontrar un resultado adecuado [Jain 2010]. En este contexto el grupo ha desarrollado FAUM (Fast Autonomous Unsupervised Multidimensional), un algoritmo de Clustering automático que puede descubrir agrupaciones naturales en big data [Curti & Wainschenker 2018]. Se basa en generar diferentes histogramas multidimensionales, definidos como Hyper-histogramas, y elegir uno para obtener la información.

El objetivo de esta línea del proyecto es continuar el desarrollo de FAUM, buscando nuevos niveles de análisis y nuevas aplicaciones.

- Estudiar en mayor profundidad los métodos empíricos propuestos en el ordenamiento de orden cero de FAUM, y buscar nuevos métodos.
- Comparar FAUM con otros algoritmos lineales de agrupamiento en términos de precisión y complejidad computacional.
- Estudiar la integración de FAUM con otros algoritmos, tanto lineales como no lineales, a fin de aprovechar las características positivas de cada método y obtener un método superador.
- Estudiar en profundidad las implicancias del uso de la distancia de Chebyshev, y la posibilidad de utilizar otras funciones de distancia.
- Estudiar un algoritmo que permita ajustar en forma automática los parámetros de FAUM para lograr diferentes objetivos de agrupamiento o clasificación.
- Estudiar la posible extensión de FAUM añadiendo un ordenamiento de nivel dos que tenga en cuenta aspectos no lineales del agrupamiento. Por ejemplo, utilizar técnicas de crecimiento local o de comparación de densidad.
- Estudiar la utilización de FAUM para identificar cubiertas de cultivo a partir de imágenes satelitales, comparando los resultados con verdades de campo.
- Estudiar la aplicación de FAUM en otros campos de estudio del aprendizaje de máquina o la minería de datos.

Sistemas Médicos de Asistencia al Diagnóstico y al Tratamiento

El objetivo de este tema es investigar métodos basados en machine learning que permitan la extracción de información a partir de la simulación del cálculo de dosis en radioterapia, el procesamiento de imágenes médicas como es el caso de las imágenes de Rayos X para la caracterización de hueso trabecular y de de imágenes de Ultrasonido Intravascular (IVUS) para la detección de estructuras vasculares, de las señales de ECG para la detección del síndrome de Bayès y de datos capturados a partir de herramientas específicamente construidas para el seguimiento y rehabilitación neurológica de pacientes. Específicamente se propone:

- Determinación de la dosis en radioterapia a través de la simulación por el método de Montecarlo para la verificación de detectores de tamaño reducido.
- Explorar diferentes técnicas basadas en Deep Learning sobre secuencias de imágenes IVUS para la caracterización de diferentes estructuras de interés sin necesidad de desarrollar algoritmos específicos para la extracción de características. Aplicar estas técnicas para la caracterización trabecular en el diagnóstico de Osteoporosis por medio de descriptores de textura.
- Aplicación de métodos de Machine Learning y específicamente de Redes Neuronales Convolucionales (CNN) para el conteo de células y proteínas en imágenes de microscopía de fluorescencia hiperespectral.
- Detección de síndrome de Bayès en señales de ECG.
- Desarrollo de herramientas de rehabilitación cognitiva para tratamiento y seguimiento de pacientes con el objetivo de extraer información a partir de la interacción del mismo con las herramientas.
- Desarrollo de monitor de postura para Camptocormia.
- Aplicación de técnicas de machine learning en problemas de clusterización y alineamiento temporal de actividades humanas a partir de videos.

Sistemas de Información Aplicados a la Cadena Productiva

Desarrollar y aplicar técnicas y algoritmos que permitan la obtención automática de información a partir de datos presentes en imágenes digitales de diferentes orígenes, que tengan como objetivo consolidar una herramienta para la toma de decisiones en las diferentes etapas de la cadena productiva de nuestro territorio.

- Continuar con el uso de imágenes del territorio, proveniente de satélites o vehículos aéreos no tripulados -drone-, en los procesos de extracción de información espacial a partir del análisis y la posible agrupación de datos similares para su integración directa en Sistemas de Información Geográfica.
- Continuar avanzando en el desarrollo de herramientas de procesamiento de imágenes que permitan asistir al desarrollo de buenas prácticas en actividades relacionadas a la agricultura y al proceso logístico agroindustrial.

Estado actual del conocimiento

A continuación se presentará el estado actual del conocimiento de cada una de las áreas de aplicación de este proyecto:

Modelado y Procesamiento de Datos Convencionales y no Convencionales

Tanto el Modelo de Entidades y Relaciones original (MER) como el Extendido (MERExt) [Chen 1976], simbolizan conceptos de datos en términos de entidades y relaciones [Elmasri & Navathe 2010] [Teorey 1990] [Teorey et al. 2011]. El minimalismo de su notación ha sido exitoso, proporcionando mecanismos de abstracción suficientes para especificar reglas del negocio y restricciones inherentes. Los lenguajes gráficos o visuales como el Diagrama de Entidades y Relaciones Extendido (DERExt) facilitan la discusión, comunicación y validación por parte de expertos en el dominio [Cuadra et al. 2013]. Muchos autores han escrito acerca de las reglas de transformación del DERExt al EFD proponiendo algoritmos con diferentes variantes [Halpin & Morgan 2010] [Teorey 1990] [Teorey et al 2011]; en particular, algunos reconocen la etapa de diseño lógico estándar, cuya definición resulta más apropiada para la etapa de transformación del DERExt a un Esquema Lógico Estándar (ELE) inicial, basado en un Modelo Objeto-Relacional. Estos algoritmos son mejorables ya que una variedad de conceptos plasmados en el DERExt se pierde al obtener el ELE/EFD. En algunos casos no es posible un proceso inverso de reconstrucción sin ambigüedades de un DERExt a partir de un ELE o de un EFD. Por lo anterior es necesario especificar un ELE que incluya todos los aspectos del modelo conceptual de datos y que extienda el algoritmo de transformación para resolver cuestiones relativas a ambigüedades y falta de ortogonalidad.

Actualmente se están desarrollando numerosas aplicaciones de software para la incorporación y el tratamiento de nuevas fuentes de datos; muchas de ellas se inspiran en desarrollos científico- tecnológicos en contextos específicos como: el estudio de las características morfológicas de una población, utilizando configuraciones de puntos anatómicos homólogos (landmarks) [Torcida et al. 2014] [Torcida et al. 2016] [Ferraggine et al. 2016], datos de curvas o de superficies; el análisis de imágenes obtenidas a partir de datos de señales acústicas generadas por vehículos autónomos submarinos; o datos de objetos digitales producto de la transformación de fuentes documentales, por ejemplo.

Dichos datos provenientes de diversas fuentes de información constituyen un campo de investigación actual y en constante evolución. Más, específicamente los Sistemas de Información Geográficos (SIG), Documentales, Morfométricos, etc. plantean nuevos desafíos, abren diferentes perspectivas y motivan el desarrollo de metodologías de procesamiento.

Técnicas Cuantitativas Orientadas al Reuso Semántico de Modelos de Requisitos

En trabajos anteriores, con el fin de detectar agrupamientos de símbolos, se aplicaron métodos dirigidos por fuerzas en la visualización de los grafos correspondientes a los Léxicos de diferentes casos de estudio. Para esta visualización, cada símbolo del LEL es representado mediante un nodo, y las menciones a otros símbolos incluidas en su definición, son representados por arcos dirigidos a los nodos respectivos. Los nodos son ubicados al azar en el marco de trabajo, y posteriormente se va modificando su ubicación en forma iterativa, mediante la aplicación de fuerzas atractivas y repulsivas.

En [Ridao & Doorn 2013] [Ridao & Doorn 2015] [Ridao & Doorn 2016] [Ridao & Doorn 2018] se presentaron los resultados de la utilización de diferentes conjuntos de fórmulas para las fuerzas aplicadas en el sistema. Se utilizaron, en primer término, las fórmulas propuestas por [Fruchterman & Reingold 1991], luego las fuerzas propuestas por [Eades 1984], y por último fórmulas propuestas por los autores de los trabajos mencionados. En cada instancia, se fueron obteniendo mejores resultados en la detección de agrupamientos. Y en todos los casos, se comprobó que los grupos detectados representaban efectivamente núcleos semánticos en los macrosistemas correspondientes.

Pero esta visualización afronta dos limitaciones relevantes. Los límites entre los grupos suelen volverse difusos, especialmente cuando los clusters poseen una forma irregular, y la variación en la granularidad con que se distribuyen las distancias entre nodos en los distintos clusters dificulta la percepción de los agrupamientos. Por ello, se pretende encontrar mecanismos para delimitar automáticamente los clusters.

Clusterización en Imágenes Multidimensionales

Los algoritmos de Clustering se pueden dividir en tres grupos generales: Jerárquicos, Basados en Partición y Basados en Grilla.

Los algoritmos Jerárquicos encuentran agrupaciones anidadas, de manera recursiva ya sea a modo de agrupación o división. Para agrupar se parte de considerar cada punto como un cluster y se empieza a agrupar de a pares por semejanza de datos. En los pasos siguientes se agrupan los pares del paso anterior, hasta que se consiga el cumplimiento de algún criterio. Por otro lado, el modo divisivo comienza con todos los puntos agrupados en un único cluster al que se lo dividirá recursivamente en clusters más pequeños hasta que se cumpla algún criterio. Ejemplos de éstas estrategias son AGNES y DIANA, Genie y DenPEHC [Gagolewski et al. 2016] [Xu et al. 2016].

Los algoritmos de clustering basados en partición encuentran todos los clusters simultáneamente realizando una partición de los datos. La partición está basada en funciones de similitud o disimilitud que no imponen una estructura jerárquica. De los algoritmos basados en partición el más popular y simple es K-Means y sus variantes [MacQueen 1967]. La complejidad de estos algoritmos es O(n*k*l) siendo “n” la cantidad de datos, “k” la cantidad de clusters y “l” la cantidad de iteraciones. Otra técnica más moderna es Affinity Propagation y sus variantes como MEAP [Wang et al. 2013] [Serdah & Ashour 2016]. La complejidad de estos algoritmos es O(N^2) [Wang et al. 2013] [Serdah & Ashour 2016] [Refianti et al. 2017] lo que los hace inconvenientes para el tratamiento de grandes bases de datos. Los algoritmos basados en grilla se diferencian de los demás métodos porque centralizan su agrupación en los valores espaciales que rodean a cada dato y no en cada dato específicamente. Dentro de esta

categoría están incluídos los métodos de densidad. Como ejemplos de éstos métodos pueden citarse STING y CLIQUE.

Un histograma parte el espacio de características en baldes (buckets) o bines de distribución uniforme. Para que el histograma describa la forma general de la curva debe ajustarse correctamente el tamaño del bin. Existen diferentes métodos basados en estadística propuestos para ajustar el tamaño del bin, como por ejemplo los descriptos por D.W. Scott, H.A. Sturges, D. Freedman y P. Diaconis y G.R. Terrell, entre otros.

Clásicamente los histogramas en espacios multidimensionales, definidos como hyper-histogramas, se estudiaron usando sus proyecciones sobre cada característica, obteniéndose un histograma por cada dimensión. Otra manera de estudiarlos era crear hyper-bines y agruparlos usando un método basado en grilla para formar los clusters.

Cálculo de Dosis en Radioterapia

La determinación de la dosis de radiación absorbida en tejido humano es de suma importancia para lograr un tratamiento de radioterapia eficaz. El método más preciso para estimar la dosis es el cálculo basado en Montecarlo pero los programas que realizan este cálculo han debido optar por alguna de las variantes en el compromiso, aún no resuelto apropiadamente, entre la calidad de la estimación y el costo temporal del cálculo. Muchas de las técnicas existentes de reducción de tiempo se basan en una simplificación del problema y acarrean una pérdida de calidad en los resultados. Se han aplicado técnicas de cálculo directo, el precálculo y la paralelización que permiten reducir el tiempo de cálculo sin perder calidad en los resultados realizando un cálculo completo sin simplificaciones. Las simulaciones en el contexto de dimensiones reducidas presentan dificultades que imponen desafíos a solucionar por parte de los programas de cálculo como PENELOPE [Salvat et al. 2008]. Existe una tendencia general al uso de haces altamente conformados que depositan dosis de gran magnitud en campos pequeños, pero en estos casos no siempre es posible asegurar que la planificación de la distribución de dosis prescrita por el oncólogo se corresponda con la entregada al paciente durante el tratamiento. Bajo tales condiciones de tratamiento se recomienda el uso de nuevos sistemas para dosimetría en tiempo real y con alta resolución espacial, para poder controlar in-vivo la calidad del tratamiento radiante [Papaconstadopoulos et al. 2014]. Recientemente ha surgido una técnica conocida como dosimetría por fibra óptica (DFO), la cual se basa en el uso de una pequeña pieza (<1 mm3) de material radioluminiscente acoplada al extremo de una fibra óptica [Andreo et al. 2015].

Procesamiento de Imágenes Médicas

En el ámbito de las imágenes médicas la exactitud en el diagnóstico y/o evaluación de una enfermedad depende tanto de la adquisición de imágenes como de la interpretación de las mismas. La primera fue mejorada sustancialmente en los últimos años, pero solo recientemente se ha empezado a ver resultados efectivos en la interpretación de esas imágenes de manera automática asistido mediante Inteligencia artificial [Greenspan et al. 2016].

La caracterización automática o semi-automática de estructuras anatómicas en imágenes IVUS es un tema de gran interés para su aplicación a Sistemas de Asistencia al Diagnóstico. Además, este problema constituye un desafío para el campo de las técnicas de visión computacional debido, entre otras cosas, al gran volumen de información que se debe procesar, el alto nivel de ruido presente en las imágenes y a la similitud de las estructuras que se deben segmentar [Katouzian et al. 2012] [Balocco et al. 2014].

Desde hace aproximadamente una década se ha comenzado a utilizar la información presente en imágenes médicas como complemento a la Densidad Mineral Ósea para el diagnóstico de Osteoporosis. Para ello, es necesario caracterizar la estructura del hueso trabecular. En este sentido, se han aplicado diferentes técnicas de procesamiento de imágenes de Rayos-X, RMN, CT y más recientemente sobre imágenes de Rayos X de doble haz (DXA) con diferente grado de éxito [Sapthagirivasan et al. 2013] [Cruz et al. 2018]. Algunas de estas técnicas [Singh et al. 2017] [Paul et al. 2017] se basan en métodos de Machine Learning y los últimos trabajos se han concentrado en el uso de CNN, aunque no aún con la finalidad de caracterización, sino como mecanismo de extracción de características.

Procesamiento de imágenes biológicas microscópicas

En el problema de cuantificación y caracterización en imágenes de microscopía de Fluorescencia, las técnicas tradicionales de procesamiento de imágenes han logrado un grado de efectividad alta, específicamente en algunos problemas [Bordacahar et al. 2016]. Sin embargo, la variabilidad de estas imágenes hace que sea necesaria la aplicación de técnicas de Machine Learning. Específicamente en el caso de imágenes multiespectrales de microscopía de fluorescencia, recientemente se ha comenzado a aplicar técnicas basadas en CNN para realizar esta tarea [Xie et al 2015] [Kraus et al. 2015].

Detección de síndrome de Bayès

En los últimos años se ha demostrado la asociación del Síndrome de Bayés a múltiples afecciones médicas del sistema circulatorio [Bayès et al. 2017]. En este sentido es relevante la detección de este síndrome a partir del estudio de la alteración en la onda P del ECG. Esto puede realizarse con técnicas de procesamiento de señales, pero al día de hoy sólo existen soluciones disponibles propietarias [Macfarlane et al. 2005].

Herramientas de Rehabilitación Cognitiva

Existen numerosas investigaciones que muestran resultados beneficiosos en la aplicación de diferentes herramientas de software para el problema de rehabilitación cognitiva en pacientes neurológicos. En este sentido hay un conjunto de herramientas, la mayoría con licencias propietarias [Rainbow 2018]. Los mayores inconvenientes de estas herramientas es que se ofrecen para un número limitado de plataformas de hardware y de

software, lo cual limita el acceso a los pacientes, además muchas de ellas ofrecen características limitadas de personalización y requieren asistencia de otra persona para el acceso al software. En cuanto a la clusterización temporal de actividades humanas a partir de videos, actualmente se está trabajando con técnicas derivadas de K-Means y Clustering Jerárquico [Zhou et al. 2013] y el método de Mallows [Sener et al. 2018].

Sistemas de Información aplicados a la cadena productiva

El desarrollo de herramientas de software basados en imágenes que asistan a diversas áreas productivas involucran por lo general varios paradigmas: manejo de algoritmos de procesamiento de imágenes para extraer información de interés y gestión de bases de datos con capacidades geográficas y SIG. Con el afianzamiento de estas áreas de investigación [Tristan et al. 2009], surge la posibilidad de desarrollar un conjunto de herramientas orientado a la web o sobre plataformas móviles, que permitan a bajo costo, resolver problemas de aplicación concretos.

Particularmente, el sector agropecuario, se ha convertido en una actividad económica con cada vez más demandas, con precios cada vez más ajustados y con mayores exigencias en cuanto a calidad de los alimentos y el respeto al ambiente. En este aspecto es donde, las condiciones de la infraestructura y la accesibilidad son elementos claves en la eficiencia de las organizaciones productivas para maximizar la competitividad del territorio [Tristan et al. 2018].

La diversidad de imágenes obtenidas desde plataformas satelitales nos permite analizar el territorio desde una perspectiva cada vez más completa. La variedad de resoluciones espaciales hace posible estudios a diferentes escalas, mientras que la información espectral posibilita la caracterización del territorio [Leguizamón et al. 2018]. Todo ello, unido a la periodicidad de adquisición de las imágenes, hace que estas técnicas sean idóneas para seguir la evolución del territorio a lo largo del tiempo.

Otro aspecto interesante de análisis, es la determinación de la calidad y clasificación de la cosecha. La calidad de los granos define las condiciones de la comercialización, por ende el valor de comercialización y un posible incremento en el ingreso de divisas al país. Actualmente la determinación de la calidad de los granos es realizada en forma manual por los peritos clasificadores de granos, los cuales se basan en Normas y Estándares Oficiales. En este sentido, los avances tecnológicos, posibilitan hoy contar herramientas de asistencia a la clasificación que permitan generar un análisis preliminar de la calidad del grano en el lugar mismo de la cosecha, pudiendo eventualmente cambiar la configuración de la máquina cosechadora con el objetivo de mejorar la calidad de la mercadería obtenida.

Siguiendo la cadena de comercialización de granos, se ha planteado la necesidad de poder detectar o clasificar semillas de diferentes granos de manera automática cuando éstas están siendo transportadas para su almacenamiento. En las grandes plantas de acopio de cereales, las semillas son transportadas entre silos mediante cintas transportadoras que se desplazan rápidamente, cualquier error humano que implique una mezcla de granos resultaría en una importante pérdida de dinero. Una detección en tiempo real de la mercadería transportada y de la no presencia de semillas de diferente variedad resultaría de gran importancia a la hora del movimiento y almacenamiento de cereales.

Metodología

En relación a lo planteado en el párrafo introductorio, para atacar el problema de la extracción de información en el contexto de Ciencia de Datos, la metodología se centra en técnicas que permiten extraer información utilizando mecanismos de aprendizaje como Machine Learning, en particular métodos No Supervisados como Clusterización y Supervisados, como aquellos basados en Deep Learning. A continuación se presentan las metodologías específicas para cada una de las áreas.

Modelado y Procesamiento de Datos Convencionales y no Convencionales:

Para estudiar los modelos de datos y el procesamiento provenientes de diversas fuentes de información como los Sistemas de Información Geográficos (SIG), Documentales, Morfométricos y otros se plantea:

- Estudio de incorporación de especificaciones formales de las transformaciones para el entendimiento de la semántica del modelo de datos plasmado en el EFD.
- Análisis de posibles representaciones de datos morfométricos en bases de datos Sql-NoSql.
- Análisis de las capacidades geométricas de los SIG actuales para representar datos de señales acústicas producidas por vehículos submarinos autónomos.
- Desarrollo de metodologías de procesamiento para los datos morfológicos y de señales acústicas mencionados.

Técnicas Cuantitativas Orientadas al Reuso Semántico de Modelos de Requisitos:

Con el fin de mejorar la visualización de los agrupamientos presentes en los modelos de IR, y analizar la información presente en los clusters detectados, se propone:

- Crear mecanismos para delimitar automáticamente los clusters. En principio, se propone utilizar como parámetro la distancia máxima entre nodos vecinos de un mismo cluster.
- Revisar los arcos emergentes de los nodos frontera de los agrupamientos detectados con el fin de descubrir símbolos faltantes, reduciendo así el nivel de omisiones en los modelos estudiados.
- Identificar y resaltar los nodos que poseen el rol de vincular agrupamientos cercanos con el fin de detectar puntos críticos en el proceso del negocio que muy posiblemente serán a su vez puntos críticos del sistema de software.

- Revisar nodos apartados o muy apartados de los agrupamientos detectados, analizando si se trata de símbolos mal incluidos o símbolos mal descriptos,
- Analizar la estructura de las relaciones del LEL del Universo del Discurso con el LEL de los requisitos para estimar la calidad del mapeo semántico de las necesidades del usuario con la lista de requisitos obtenida.

Clusterización en Imágenes Multidimensionales

Para hallar los clusters en un espacio de datos multidimensional de manera computacionalmente eficiente, FAUM trata el proceso como una sucesión de pasos. Cada paso extrae información relevante de los datos de entrada, generando otro conjunto de datos más pequeño que se usará en el paso siguiente. El primer paso, que trabaja con bases de datos extensas, tiene la menor complejidad computacional posible. Por otro lado, los pasos siguientes, que trabajan con conjuntos de datos menores, podrán tener mayor complejidad computacional. En el trabajo ya publicado [Curti & Wainschenker 2018] se han propuesto dos pasos. En el primer paso se construyen hiper-histogramas usando cierto criterio para determinar un tamaño de hiper-bin adecuado. A este paso se lo ha llamado agrupamiento de orden cero. n el segundo paso, denominado agrupamiento de orden uno, se hallan clusters agrupando hyper-bines por proximidad. Para el agrupamiento de orden uno es importante definir un criterio de proximidad y elegir una adecuada función de distancia, considerando la naturaleza multidimensional del espacio [MacQueen 1967].

Cálculo de dosis en Radioterapia

En radioluminiscencia, la medición de la intensidad de luz transmitida por un centellador a través de una fibra óptica sirve para determinar la tasa de dosis en tiempo real e in-vivo. La pequeñez del centellador confiere alta resolución espacial al sistema dosimétrico [Kim et al. 2012]. El desarrollo de esta metodología puede verse beneficiado por la realización de simulaciones que permitan analizar el comportamiento de tales detectores, respaldando y explicando los resultados obtenidos con la utilización de detectores de campo pequeño. La simulación de campos pequeños presenta desafíos provocados principalmente por deficiencias en las estadìsticas de partículas en los detectores. Esta deficiencia provoca un aumento del error en los resultados y en este sentido se está trabajando en realizar simulaciones con caracterìsticas especiales para solucionar los problemas estadìsticos [Martinez et al. 2017]. De esta forma, pueden ser analizadas las mediciones realizadas in-situ en condiciones clínicas de tratamientos de campos pequeños, en conjunto con el respaldo provisto por las metodologías de simulación para la interpretación de tales resultados [Massa et al. 2007] [Massa et al. 2010]. La metodología a aplicar respecto a los problemas de simulación de campo pequeño se centra principalmente en el desarrollo de herramientas que agilicen las simulaciones actuales y mejoren el análisis de sensibilidad de los parámetros de simulación respecto de la calidad de los resultados. Se planifica adecuar el programa PENELOPE para el contexto de dimensiones reducidas. Se considera además, el uso de técnicas de reducción de varianza específicas como propuesta para la solución de problemas de campo pequeño. Con el objeto de agilizar la planificación de las simulaciones, se propone el desarrollo de una herramienta que permita asistir al diseño de las geometrías 3D que permita a los investigadores centrarse en la información específica de la simulación, facilitando su inicialización y optimizando el acceso a la información relativa a los materiales involucrados. De manera adicional, se propone el estudio de la aplicación y adaptación de las herramientas disponibles para la simulación para su empleo en el contexto de detectores de campos pequeños.

Procesamiento de imágenes médicas

Particularmente desde hace algunas décadas se ha comenzado a utilizar la metodología denominada IVUS como complemento indiscutible para el procedimiento denominado Angioplastia con fines de caracterización y cuantificación de la placa ateroesclerótica. Esta técnica permite obtener una secuencia de imágenes ecográficas en tiempo real y de alta resolución del interior de los vasos sanguíneos de los pacientes. Sin duda la segmentación manual de cualquier patrón de interés dentro de este tipo de imágenes es una tarea laboriosa, lo que constituye un reto para su segmentación automática. En los últimos años se han publicado varios trabajos que utilizan técnicas de AP para la detección de patrones de interés en imágenes médicas, algunos de ellos muy interesantes como el presentado por [Cernazanu-Glavan & Holban 2013] una CNN capaz de segmentar la estructura de los huesos en imágenes de rayos X. En [de Brebisson & Montana 2015] se aplica un enfoque similar pero en imágenes cerebrales tipo MRI (Magnetic Resonance Imaging) para realizar una clasificación en 3 Dimensiones de los voxeles que corresponden a cada región anatómica. Recientemente en esa misma área de la medicina [Havaei et al. 2017] presentaron un método de segmentación automática de tumores cerebrales basado en CNNs.

Especialmente en este área, las técnicas de Machine Learning se encuentran con desafíos tales como el manejo de grandes volúmenes de datos, la existencia de diferentes modalidades de imágenes, la falta de información etiquetada por expertos y otros defectos propios de las imágenes médicas. Luego de analizar la bibliografía del área de estudio, no se han encontrado trabajos donde se apliquen técnicas de AP sobre imágenes IVUS.

En el caso de la caracterización trabecular, las técnicas de análisis de textura son adecuadas para obtener un indicador de la estructura ósea. Si bien estas técnicas han resultado efectivas sobre imágenes de Rayos-X, CT y RMN, presentan un desafío sobre las imágenes de DXA debido entre otras causas a su baja resolución. Para ello, se propone emplear técnicas de Machine Learning y específicamente CNN para la caracterización trabecular.

Procesamiento de imágenes biológicas microscópicas

En el análisis de objetos en imágenes de microscopía de alta resolución se presentan algunos desafíos relacionados con la cantidad de información a procesar y los problemas propios de los experimentos, como las variaciones en las condiciones de iluminación, en el ángulo de captura de la imagen, resolución y superposiciones,

entre otros. Se propone atacar el problema del conteo de células en imágenes de fluorescencia por medio de técnicas de Deep Learning a través de diversos enfoques basados en propuestas de regiones, ventaneo y técnicas mixtas en las que se aplica un algoritmo tradicional para extraer las características y luego una CNN para clasificar de acuerdo a estos vectores característicos.

Detección de síndrome de Bayès

El desarrollo de un método de detección que pueda resultar útil para el diagnóstico del Síndrome de Bayès, supone la creación de un dataset apropiado. En este sentido, se está trabajando con la base de datos aportada por el propio Antoni Bayès de Luna, lo cual implica realizar trabajos de digitalización de ECG que preserven la morfología de la onda P, la cual es vital en el diagnóstico de esta patología.

Herramientas de Rehabilitación Cognitiva:

Se plantea la utilización de metodologías dirigidas por el modelo y tecnologías multiplataforma a través del uso de lenguajes y transcompiladores como por ejemplo Haxe [Ponticelli 2008] que permitan la implementación del software en un lenguaje de alto nivel que pueda ser traducido a múltiples plataformas manteniendo un único código fuente, lo cual favorece la mantenibilidad, escalabilidad y seguridad entre otros aspectos. Con el fin de agregar funcionalidad de diagnóstico, se propone la utilización de técnicas de detección de gestos y de mirada que permitan capturar con mayor profundidad y detalle el uso de la herramienta por parte de los pacientes. Con el fin de clasificar estas interacciones, se propone el uso de técnicas de clustering temporal a partir de videos.

Sistemas de Información aplicados a la cadena productiva

Con el fin de afianzar el uso de información del territorio extraída desde sensores remotos para el desarrollo de la cadena productiva, se propone continuar profundizando el conocimiento de las técnicas de extracción de información y las posibles alternativas de mejora así como incursionar en nuevas tecnologías de clasificación basadas en Deep Learning y su aplicación a problemas que requieren detección en tiempo real.

Plan de actividades totales, estado de avance de la línea y cronograma

Si bien el dominio de aplicación de cada área tiene particularidades propias, en muchos casos se comparten las características de los datos mencionadas en la introducción. Por ejemplo, la característica del volumen de los datos está presente en varios de los dominios estudiados en este proyecto, como el de las imágenes biológicas multiespectrales y el de las imágenes satelitales de múltiples bandas, entre otros. Entonces, dado que las características afectan en gran medida la técnica o método a aplicar sobre los datos en cuestión, se planifica trabajar en forma colaborativa en el desarrollo y aplicación de dichas técnicas. De esta manera, será posible extrapolar el uso de una técnica aplicada en un área a otra. A continuación se presenta para cada área, el plan de actividades.

Modelado y Procesamiento de Datos Convencionales y no Convencionales

Se ha propuesto un algoritmo de transformación que permita obtener un ELE sin pérdida de conceptos plasmados en el modelo conceptual. Se ha finalizado el desarrollo de la herramienta CASE que sigue ciclo de diseño de una base de datos desde el modelo conceptual. Se han obtenido de manera automática scripts compilables de SQL4. Se planifica:

- Incorporar en las transformaciones aspectos relativos a los vínculos semánticos ternarios. (6 meses)
- Evaluar la utilización de Model-Driven engineering (MDE) en las transformaciones de modelos para permitir su automatización. (6 meses)
- Aplicar la metodología mencionada en un nivel superior de desarrollo, aplicando los conceptos en los casos de big data. (12 meses)
- Estudiar las características comunes y diferenciales que permitan el tratamiento de los datos no convencionales especificados. (12 meses)

Técnicas Cuantitativas Orientadas al Reuso Semántico de Modelos de Requisitos

Durante el primer año del proyecto, se trabajará en la propuesta de mecanismos para la detección automática de clusters presentes en el LEL. Una vez logrados los resultados esperados, estos mecanismos serán aplicados al resto de los modelos del proceso de IR, lo que se llevará a cabo a lo largo del resto del período. En paralelo con estas actividades, se continuará con la aplicación de técnicas cuantitativas y/o cualitativas a todos los modelos del proceso, con el fin de obtener información no visible directamente en ellos.

Clusterización en Imágenes Multidimensionales

La primera versión de FAUM, ya desarrollada y publicada, propone dos pasos de agrupamiento. Con estos dos pasos FAUM es una algoritmo de agrupamiento lineal de simetría esférica. Desde este punto surgen varias líneas de trabajo que pueden desarrollarse en forma paralela. Por un lado se estudiará la aplicación de FAUM, en su estado actual de desarrollo, a diferentes problemas actuales, como ser la identificación de cubiertas de cultivo a partir de imágenes satelitales. Por otro lado se estudiarán en mayor profundidad las decisiones tomadas y los métodos propuestos en busca de optimizaciones. Finalmente, por un tercer camino, se estudiarán las extensiones posibles de FAUM combinándolo con otras técnicas en la forma de nuevos órdenes de agrupamiento. Se espera que FAUM pueda entonces comportarse como un algoritmo no lineal [Wang & Lai 2016] y por lo tanto reconocer agrupamientos mucho más complejos.

Sistemas médicos de asistencia al diagnóstico y al tratamiento

En el problema de simulación de detectores pequeños se ha trabajado en la modificación de espacios de fase

existentes para adecuarlos al contexto. Se ha estudiado la dependencia angular de la dosis en detectores de tamaño reducido. Se planifica continuar con este trabajo en el marco de la tesis de doctorado en progreso del Lic. Nahuel Martínez cuya beca es co-dirigida por el Dr. José M. Massa con la colaboración de investigadores del Instituto de Física Arroyo Seco y del Arbeitsgruppe Strahlungsphysik de la TU-Dresden, Alemania.

Respecto de la caracterización trabecular, se ha retomado un trabajo previo [Pecelis, 2009] para su aplicación sobre imágenes de DXA. En este momento se está evaluando la eficacia del método vectorial publicado sobre estas imágenes, para lo cual se firmó un convenio con el Instituto de Diagnóstico De Martino de la ciudad de Tandil. Cronograma: Noviembre de 2017: Extracción de imágenes en formato RAW de equipo de DXA (realizado). Junio de 2018: Aplicación del método vectorial sobre imágenes de DXA (realizado). Diciembre de 2018: Aplicación de técnicas de correlación sobre imágenes de DXA (a realizarse). Enero a Diciembre de 2019: Estudio comparativo de estas técnicas entre imágenes de DXA y Rayos-X (a realizarse).

En relación al procesamiento de imágenes biológicas, se comenzaron a aplicar técnicas tradicionales y basadas en Deep Learning sobre conteo de células, proteínas y neuronas en imágenes de microscopía de fluorescencia. Se ha presentado un trabajo preliminar en un workshop internacional, el cual se planifica extender en un artículo conteniendo una comparativa de diferentes enfoques basados en CNN. En este sentido se está comenzando a trabajar en colaboración con el Área de Ciencias Morfológicas de la Facultad de Ciencias Veterinarias, UNICEN y con el Instituto de Medicina y Biología Experimental de Cuyo (IMBECU) para la realización de capacitaciones conjuntas. Cronograma: Diciembre de 2017: Desarrollo de herramienta basada en detección de objetos para el conteo de núcleos. Julio de 2018: Implementación de técnicas de CNN basadas en propuestas de regiones, ventaneo y técnicas mixtas. Enero 2019 a Julio de 2021. Entrenamiento de CNN a gran escala y desarrollo de métodos de identificación de proteínas.

En cuanto a la detección del Síndrome de Bayès, durante los años 2017 y 2018 se ha contactado con el equipo del Dr. Antoni Bayès y se ha comenzado a trabajar en la digitalización del conjunto de ECG aportado por el equipo. Se planifica desde el año 2019 hasta el año 2021 continuar con la tesis de Doctorado de la Lic. Lorena Franco quien trabajará en métodos de detección de Síndrome de Bayès bajo la dirección del Dr. José M. Massa [Franco, 2018].

En lo que se refiere al desarrollo de Herramientas de Rehabilitación Cognitiva, se ha realizado un convenio marco aprobado por resolución RJE 4751-12 entre la UNICEN y el Hospital Italiano de Buenos Aires (HIBA) y se ha firmado y finalizado un convenio específico aprobado por resolución RR1097/16 en los que se desarrolló un módulo de la herramienta de rehabilitación. En el transcurso del año 2018 se está trabajando en un nuevo convenio específico para el módulo de rehabilitación de memoria y se prevé en los próximos años implementar dos módulos de rehabilitación motora y cognitiva. Durante el año 2018 se han implementado técnicas de alineamiento temporal en secuencias de videos de actividades humanas y se planifica analizar su implementación en el contexto de las herramientas de rehabilitación cognitiva durante el período 2019-2021.

Sistemas de Información aplicados a la cadena productiva:

El plan de actividades propuesto continúa desarrollos preexistentes:

- Avanzar en la mejora de técnicas para extracción automática de información provenientes de imágenes digitales.
- Profundizar el estudio de nuevas tecnologías de clasificación no supervisada.
- Implementar técnicas que den soporte a la detección en tiempo real de granos durante el movimiento y almacenaje.
- Promover la utilización de la información proveniente de imágenes en procesos de planificación para la mejora de la competitividad de nuestro territorio.

Aportes académicos y de transferencia esperados

Los conocimientos adquiridos en las diferentes áreas se volcarán en publicaciones, actividades de docencia, realización de postgrados, proyectos de extensión, proyectos de transferencia, etc.:

Publicaciones de trabajos de investigación

Se espera que los resultados de las investigaciones en las diferentes áreas sean publicados en eventos tanto nacionales como internacionales y revistas especializadas en los temas respectivos.

Se continuará con la redacción de un libro de texto didáctico para el tema Lenguajes de Programación que se viene desarrollando desde el proyecto anterior.

Actividades de docencia

Los diferentes temas del presente proyecto han dado lugar a diferentes proyectos de Trabajo Final de las carreras de la Facultad de Cs. Exactas. Se espera continuar con esta modalidad.

Los trabajos del área de Modelado y Procesamiento de Datos Convencionales y no Convencionales se vuelcan en el dictado y confección de material didáctico para el área de Bases de Datos y Estructuras de Almacenamiento, de la carrera Ing.de Sistemas y en Introducción a las Bases de Datos y Bases de Datos de las Tecnicaturas TUPAR y TUDAI, de la Fac. de Cs.Exactas de UNCPBA. Los resultados del área Técnicas Cuantitativas Orientadas al Reuso Semántico de Modelos de Requisitos se vuelcan en el dictado y confección de material didáctico para materias optativas (Ing. de Requisitos), y el curso de posgrado Tópicos de Ingeniería de Requisitos de UNLaM.

Se ha planificado el dictado de un curso de Procesamiento de Imágenes Biológicas para investigadores del área de Biología de diferentes Institutos Nacionales.

Se espera continuar el dictado del curso de Introducción a los métodos de Monte Carlo para el doctorado en Física con la colaboración del Dr. Cravero de la Universidad Nacional del Sur.

Se planifica el dictado de un curso “Taller de Lenguajes de Programación” en base al material generado por la redacción en curso de un texto de estudio de Lenguajes de Programación.

Trabajos de Investigación en Colaboración

Se trabaja en colaboración con el Grupo de Evolución Morfológica de la Fc.de Cs.Nat. y Museo de la UNLP, en el desarrollo de aplicaciones de software para diferentes análisis morfométricos y se han realizado presentaciones conjuntas en congresos y revistas del área.

Se prevé colaboración con proyectos de Investigación de la Universidad Nacional de La Matanza, así como la Universidad Nacional del Oeste.

Se han iniciado trabajos en colaboración con INTA Pergamino en la aplicación de FAUM para la identificación de cubiertas de cultivos a partir de imágenes satelitales.

Se planifica continuar la colaboración con el Arbeitsgruppe Strahlungsphysik de la Technische Universität Dresden, el laboratorio de Computación Perceptual de la Universität de Barcelona y diferentes grupos de la Universidad Nacional de Sur, entre otros.

Trabajos de Transferencia y Extensión

Se trabaja en conjunto con el Instituto Nacional de Tecnología Agropecuaria INTA, en el desarrollo de aplicaciones de software orientadas a promover el desarrollo agropecuario. Se espera continuar trabajando en este sentido y que el resultado de esta investigación pueda ser transferido a éste y otros organismos públicos o privados.

Se espera poder continuar con el convenio de colaboración con el Instituto de Diagnóstico De Martino y el Hospital Italiano de Buenos Aires para la provisión de imágenes y asesoría médica.

Se espera que el resultado de la investigación de Clusterización en Imágenes Multidimensionales se puedan transferir a diferentes organismos públicos o privados.

Realización de Postgrados

Los aspectos relacionados con el Modelado y Procesamiento de Datos Convencionales y no Convencionales forman parte del plan de trabajo de Viviana Ferraggine para la tesis del doctorado en Matemática Computacional e Industrial de la Universidad Nacional del Centro de la Pcia. de Bs.As.

Las actividades propuestas para las Técnicas Cuantitativas Orientadas al Reuso Semántico de Modelos de Requisitos forman parte del trabajo de tesis de Marcela Ridao para el doctorado en Cs Informáticas de la UNLP: Técnicas Cuantitativas Orientadas al Reuso Semántico de Modelos de Requisitos.

Los trabajos del área de Sistemas médicos de asistencia al diagnóstico y al tratamiento se volcarán en la finalización de la tesis de maestría en Ing. de Sistemas del Ing. José Marone referida al área Procesamiento de Imágenes Médicas, en el avance del doctorado en Matemática Computacional e Industrial de la Lic. Franco en cuanto a la detección de síndrome de Bayès sobre imágenes aportadas por el Hospital Italiano de Buenos Aires, la finalización de la tesis de Doctorado en Física del Lic. Nahuel Martínez, contribuir a la tesis de doctorado en Matemática Computacional e Industrial del Ing. Menchón con la aplicación de técnicas de clustering y Machine Learning aplicados a los datos capturados del uso de la herramienta (dependiendo de los tiempos de puesta a disposición de la herramienta a los pacientes por parte del HIBA).

Antecedentes del grupo en la temática

La mayoría de las líneas de trabajo presentadas en las secciones anteriores tienen origen en varios Proyectos de Incentivos presentados anteriormente bajo el nombre “Bases de Datos y Procesamiento de Señales”. Cabe mencionar que este proyecto se llevará a cabo en el Instituto de Investigación en Tecnología Informática Avanzada (INTIA), el cual recientemente comenzó a formar parte del conjunto de Centros de Investigación Asociados de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires. Los temas de este proyecto se han presentado en reuniones organizadas por dicha Comisión. A continuación se detallan específicamente algunos antecedentes particulares:

- FAUM [Curti & Wainschenker 2018] el algoritmo en el que se basa esta línea de trabajo, fue creado desde el principio en este grupo.
- Respecto del cálculo de Dosis en Radioterapia, en el año 2012 se ha completado el doctorado de José Massa bajo la dirección del Dr. Rubén Wainschenker, se ha participado en diversos proyectos con el Instituto de Física Arroyo Seco, se han realizado estadías conjuntas de investigación con el grupo ASP de la TU-Dresden y se han publicado los avances en este tema en congresos y revistas especializadas [Massa et al. 2007], [Massa et al. 2010], [Martínez et al. 2017] [Martínez et al. 2018].
- En el tema de segmentación de IVUS, el Ing. Marone ha realizado una estadía de investigación en el Centro de Visión Computacional, dependiente de la UB, Barcelona, donde ha completado parte de su doctorado y finalizada una publicación en el tema [Marone et al. 2016]. En cuanto al análisis de imágenes trabeculares, se han publicado diferentes trabajos sobre técnicas clásicas para la caracterización trabecular [Iglesias et al. 2010] [Santiago et al. 2009], [Pecelis et al 2009].
- En el Procesamiento de imágenes biológicas microscópicas se ha trabajado en varios problemas de imágenes biológicas de microscopía como la identificación de cromosomas, el conteo de células epiteliales, el conteo de

células neuronales, la cuantificación de fibras de colágeno [Menchón, 2018] [Bordacahar et al 2016] [Herrera et al. 2018].

- En cuanto a la detección de síndrome de Bayès, el grupo tiene amplia experiencia en procesamiento de Señales y específicamente se han aplicado métodos de procesamiento de imágenes sobre problemas de cardiología [Claret et al. 2015].
- Con respecto al desarrollo de herramientas de rehabilitación cognitiva, uno de los integrantes del proyecto ha realizado su tesis de grado en este tema, además varios integrantes forman parte del equipo técnico en los convenios de colaboración con el HIBA. El Ing. Menchón ha realizado una estadía de investigación en la Universitat de Barcelona para especializarse en técnicas de clustering temporal. Se han publicado los avances de los trabajos en [Menchón et al. 2016-1] y [Menchón et al. 2016-2].
- En cuanto a los Sistemas de Información aplicados a la cadena productiva, con respecto al desarrollo de herramientas de clasificación y estimación de calidad de granos y desde hace algunos años se han concretado varios trabajos de tesis de grado con importantes aportes de desarrollo en la temática. Así mismo se está trabajando conjuntamente con el Grupo de Investigación en Bioarqueología de la UNICEN y un grupo de becarios de la Comisión de Investigaciones Científicas en la elaboración de una jerarquización de la infraestructura vial de la Pcia. de Buenos Aires como una herramienta de toma de decisiones para desarrollo logístico-productivo, mediante la caracterización del territorio basado en sensado remoto.

Referencias y principal bibliografía

- Aiello, A., & Silveira, A. (2004). Trazado de grafos mediante métodos dirigidos por fuerzas: revisión del estado del arte (Doctoral dissertation, Tesis de Licenciatura en Ciencias de la Computación).
- Álvarez M., Tristán P., Massa J. & Wainschenker R. (2011): Clasificación Automática de Cubiertas Terrestres en Imágenes Satelitales.XVII Congreso Argentino de Ciencias de la Computación. (CACIC). La PLata, Argentina
- Andreo, P., Palmans, H., Marteinsdóttir, M., Benmakhlouf, H., & Carlsson-Tedgren, Å. (2015). On the Monte Carlo simulation of small-field micro-diamond detectors for megavoltage photon dosimetry. Physics in Medicine & Biology, 61(1), L1.
- Badia, A., & Lemire, D.(2011): A call to arms: revisiting database design. ACM SIGMOD Record, 40(3), 61--69.
- Balocco, S., Gatta, C., Ciompi, F., Wahle, A., Radeva, P., Carlier, S. & Kovarnik, T. (2014). Standardized evaluation methodology and reference database for evaluating IVUS image segmentation. Computerized medical imaging and graphics, 38(2), 70-90.
- Bayès de Luna, A, Baranchuk, A., Robledo, L. A. E., van Roessel, A. M., & Martínez-Sellés, M. (2017). Diagnosis of interatrial block. Journal of geriatric cardiology: JGC, 14(3), 161.
- Berkhin, P. (2006). A survey of clustering data mining techniques. In Grouping multidimensional data (pp. 25-71). Springer, Berlin, Heidelberg
- Bordacahar G., Casalini, B. & Massa, J. (2016). Identifying cell protein expression in fluorescent microscopy images, Simposio. SIPAIM 2016 : 12th International Symposium on Medical Information Processing and Analysis.
- Cernazanu-Glavan, C., & Holban, S. (2013). Segmentation of bone structure in X-ray images using convolutional neural network. Adv. Electr. Comput. Eng, 13(1), 87-94..
- Chen, P. P. S. (1976): The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems (TODS), 1(1), 9--36.
- Claret, G., Gabiola, A., Díaz, A., Lo Vercio, L., & Massa, J. (2015). Semi-automatic Carotid IMT measurement using an active contours and texture approach. International Conference on Machine Vision, Barcelona.
- Codd, E. F.(1990): The relational model for database management: version 2. Addison-Wesley Longman Publishing Co., Inc.
- Cruz, A. S., Lins, H. C., Medeiros, R. V., José Filho, M. F., & Silva, S. G. (2018). Artificial intelligence on the identification of risk groups for osteoporosis, a general review. Biomedical engineering online, 17(1), 12
- Cuadra, D., Martínez, P., Castro & E. Al-Jumaily, H.(2013): Guidelines for representing complex cardinality constraints in binary and ternary relationships, Software & Systems Modeling Journal, Vol.12, 4, 871-889.
- Curti H.J., Wainschenker, R.S.(2018): FAUM: Fast Autonomous Unsupervised Multidimensional classification. Information Sciences 462 182–203.https://doi.org/10.1016/j.ins.2018.06.008
- de Brebisson, A., & Montana, G. (2015). Deep neural networks for anatomical brain segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 20-28).
- do Prado Leite, J. C. S., Doorn, J. H., Kaplan, G. N., Hadad, G. D., & Ridao, M. N. (2004). Defining System Context using Scenarios. In Perspectives on Software Requirements (pp. 169-199). Springer, Boston, MA.
- Eades, P. (1984). A heuristic for graph drawing. Congressus numerantium, 42, 149-160.
- Elmasri, R., & Navathe, S.(2010): Fundamentals of Database Systems, 6th Ed. Adison-Wesley Pub. Comp..
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd (Vol. 96, No. 34, pp. 226-231).
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Measurement of proximity. Cluster Analysis, 43-69.
- Ferraggine, V. E., Torcida, S., & Perez, S. I. (2016): RPS (Resistant Procrustes Software): Una Herramienta Novedosa para el Análisis Morfométrico Resistente. IV Congreso Nacional de Ingeniería en Informática/Sistemas de Información. UCASAL.
- Franco, L.,, Escobar Robledo L., Bayés de Luna, A. & Massa, J.; (2018) Digitalización de Imágenes de ECG para la Detección del Síndrome de Bayés, aceptado para su publicación en CACIC 2018.
- Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. Software: Practice and experience, 21(11), 1129-1164.
- Gagolewski, M., Bartoszuk, M., & Cena, A. (2016). Genie: A new, fast, and outlier-resistant hierarchical clustering algorithm. Information Sciences, 363, 8-23.

- Greenspan, H., Van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging, 35(5), 1153-1159.
- Halpin, T., & Morgan, T.(2010): Information modeling and relational databases. Morgan Kaufmann.
- Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. Medical image analysis, 35, 18-31.
- Herrera, M., Herrera, J., Massa, J., Aguilar, J. & Fumuso, E. (2018): Colágeno periglandular en endometrio de yeguas susceptibles y resistentes a endometritis, Aceptado para su publicación en el XX Congreso de Ciencias Morfológicas, La Plata, Buenos Aires, Argentina, Octubre 2018.
- Iglesias, A. F., Saleres, A. S., Vogel, W. D., Massa, J. M., Tristán, P., & Santiago, M. (2010). Técnicas de segmentación semi-automática en imágenes para diagnóstico de osteoporosis. In XII Workshop de Investigadores en Ciencias de la Computación.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666.
- Kaplan, G., Doorn, J. H., & Gigante, N. (2013). Evolución semántica de glosarios en los procesos de requisitos. In XVIII Congreso Argentino de Ciencias de la Computación. Mar del Plata, Argentina.
- Katouzian, A., Angelini, E. D., Carlier, S. G., Suri, J. S., Navab, N., & Laine, A. F. (2012). A state-of-the-art review on segmentation algorithms in intravascular ultrasound (IVUS) images. IEEE Transactions on Information Technology in Biomedicine, 16(5), 823-834.
- Kim, K. A., Yoo, W. J., Jang, K. W., Moon, J., Han, K. T., Jeon, D., & Lee, B. (2012). Development of a fibre-optic dosemeter to measure the skin dose and percentage depth dose in the build-up region of therapeutic photon beams. Radiation protection dosimetry, 153(3), 294-299.
- Kobourov, S. G. (2012). Spring embedders and force directed graph drawing algorithms. arXiv preprint arXiv:1201.3011.
- Kraus, O. Z., Jimmy, L., & Frey, B. (2015). Classification and segmenting microscopy images using convolutional multiple instance learning. arXiv preprint. arXiv preprint arXiv:1511.05286.
- Kriegel, H. P., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3), 231-240.
- Leguizamon G., Zuby J., & Tristan P (2018): Diagnostic of causes and hydrological effects in the transport infrastructure at southeast of Buenos Aires Province. Enviado para su evaluación a IV Congreso de Logística y Puertos. Puerto Quequén: Ordenamiento vehicular. Revista Técnica Especializada “Enfasis Logístico”Año XXIV Nro. 6.
- Macfarlane, P. W., Devine, B., & Clark, E. (2005, September). The university of Glasgow (Uni-G) ECG analysis program. In Computers in Cardiology, 2005 (pp. 451-454). IEEE.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 14, pp. 281-297).
- Marone, J., Balocco, S., Bolaños, M., Massa, J., & Radeva, P. (2016) Learning the Lumen Border using a Convolutional Neural Networks classifier. CVII-STENT workshop MICCAI 2016, Atenas.
- Márquez, L.; Ferraggine, V.; Perez, S. I. y Torcida, S.(2015): RPS (Resistant Procustes Software) Una Herramienta para el Análisis Morfológico. III Encuentro de Morfometría, Instituto Nacional de Limnología (INALI-CONICET-UNL), Asociación de Ciencias Naturales del Litoral (ACNL), Santa Fe, Argentina.
- Martinez N.; Fernandez Y.; Santiago M.; Molina P.; Massa J.; Marcazzó. MC study of size effects and angular response of fiberoptic scintillator detectors for radiation therapy. Argentina. Cordoba. 2018. Revista. Resumen. Conferencia. 14th International Symposium on Radiation Physics. CEPROCOR
- Martínez, N., Fernández, Y., Machiello, S., Santiago M., Molina P., Cravero W., Massa J. (2017) Study of angular dependence in Fiber Optic Dosimetry by Monte Carlo simulations. Cuba. La Habana. Revista. Resumen. Simposio. Latin-American Symposium on Nuclear Physics and Applications. NURT
- Massa, J. M., Doorn, J. H., & Wainschenker, R. S. (2010, August). Low-coupled parallel strategy for Monte Carlo radiation dose calculation. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE (pp. 1771-1774). IEEE.
- Massa, J. M., Wainschenker, R. S., Doorn, J. H., & Caselli, E. E. (2007). Time improvement of photoelectric effect calculation for absorbed dose estimation. In Journal of Physics: Conference Series (Vol. 90, No. 1, p. 012048). IOP Publishing.
- Menchón, M., Fujii, D., Rizzalli, A., Beltracchi, R., Bordacahar, G., Casalini, B. & Massa, J. (2018) Comparison of Deep Learning cell counting alternatives in Multispectral High Resolution Fluorescence images. Machine Learning Summer School. Universidad Di Tella. Argentina. Buenos Aires.
- Menchón, M., Massa, J., Marone, J. (2016), Bonifacio, A. & Golimstok, A. (2016): YART: An interactive cognitive rehabilitation tool. Argentina. Tandil. SIPAIM. 12th International Symposium on Medical Information Processing and Analysis. UNICEN
- Menchón, M., Massa, J., Marone, J. (2016). YART: Una herramienta interactiva para rehabilitación cognitiva. Argentina. CABA. 2016. Revista. Artículo Completo. Jornada. Anales de las 45 JAIIO - ISSN 1850-2776. SADIO
- Papaconstadopoulos, P., Tessier, F., & Seuntjens, J. (2014). On the correction, perturbation and modification of small field detectors in relative dosimetry. Physics in Medicine & Biology, 59(19), 5937.
- Paul, R., Alahamri, S., Malla, S., & Quadri, G. J. (2017). Make Your Bone Great Again: A study on Osteoporosis Classification. arXiv preprint arXiv:1707.05385.
- Pecelis, M., Massa, J., Velo, L. F., Santiago, M., & Caselli, E. (2009). Classification of trabecular patterns in the proximal femur using the vector representation algorithm: its correlation to the degree of osteoporosis. In World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany (pp. 1022-1024). Springer, Berlin, Heidelberg.
- Pieris, D.(2013)-1: Modifying the Entity relationship modelling notation: towards high quality relational databases from better notated ER models. Preprint arXiv:1306.5690.
- Pieris, D.(2013)-2: A novel ER model to relational model transformation algorithm for semantically clear high quality database design. arXiv preprint arXiv:1306.6734, (2013).
- Pieris, D.(2013)-3: Extending the ER Model to relational Model novel transformation Algorithm: transforming relationship Types among Subtypes. arXiv preprint arXiv:1307.4519.

- Ponticelli, F., McColl-Sylveste, L. (2008). Professional Haxe and Neko. John Wiley & Sons Inc.
- Rainbow Rehabilitation Centers@. Computer-assisted Cognitive Retraining. Recuperado de: https://www.rainbowrehab.com/computer-assisted-cognitive-retraining/
- Refianti, R., Mutiara, A. B., & Gunawan, S. (2017). TIME COMPLEXITY COMPARISON BETWEEN AFFINITY PROPAGATION ALGORITHMS. Journal of Theoretical & Applied Information Technology, 95(7).
- Ridao, M., & Doorn, J. H. (2013, June). Semántica oculta en modelos de requisitos. In XV Workshop de Investigadores en Ciencias de la Computación. Paraná, Argentina.
- Ridao, M., & Doorn, J. H. (2018). Displaying Hidden Information in Glossaries. In Encyclopedia of Information Science and Technology, Fourth Edition (pp. 7411-7421). IGI Global.
- Ridao, M., Doorn, J. (2015). Agrupamientos en Glosarios del Universo de Discurso. Tecnología y Ciencia - Revista dela Universidad Tecnológica Nacional, Edición Especial CoNaIISI 2014, 13(27), 5-16.
- Ridao, M., Doorn, J. (2016). Visualización de Núcleos Semánticos en Glosarios del Universo de Discurso. In CONAIISI 2016. Salta, Argentina, 2016.
- Salvat, F., Fernández-Varea, J. M., & Sempau, J. (2008, June). PENELOPE-2008: A code system for Monte Carlo simulation of electron and photon transport. In the Workshop Proceedings, June.
- Santiago, M. A., del Fresno, M., Massa, J. M., Escobar, P., Pecelis, M., Velo, L. F., & Caselli, E. (2009) Análisis de radiografías de fémur mediante transformada Wavelet para la detección temprana de la osteoporosis. Congreso Argentino de Bioingeniería (SABI), 1(1), 1.
- Sapthagirivasan, V., & Anburajan, M. (2013). Diagnosis of osteoporosis by extraction of trabecular features from hip radiographs using support vector machine: An investigation panorama with DXA. Computers in biology and medicine, 43(11), 1910-1919.
- Sener, F., & Yao, A. (2018). Unsupervised Learning and Segmentation of Complex Activities from Video. arXiv preprint arXiv:1803.09490.
- Serdah, A. M., & Ashour, W. M. (2016). Clustering large-scale data based on modified affinity propagation algorithm. Journal of Artificial Intelligence and Soft Computing Research, 6(1), 23-33.
- Singh, A., Dutta, M. K., Jennane, R., & Lespessailles, E. (2017). Classification of the trabecular bone structure of osteoporotic patients using machine vision. Computers in biology and medicine, 91, 148-158.
- Teorey, T. J., Lightstone, S. S., Nadeau, T., Jagadish, H. V. (2011): Database modeling and design: logical design. Elsevier.
- Teorey, T.J. (1990): Database Modeling and Design. The Entity-Relationship Approach, Morgan Kaufmann, San Mateo, CA.
- Torcida S., Gonzalez P. N. & Lotto F. (2016): A resistant method for landmark-based analysis of individual asymmetry in two dimensions. Quantitative Biology 4(4): 270–282 [DOI: 10.1007/s40484-016-0086-x].
- Torcida S., Perez S. I. & Gonzalez P. N (2014): An Integrated Approach for Landmark-Based Resistant Shape Analysis in 3D Evolutionary Biology 41(2):351-366 [DOI: 10.1007/s11692-013-9264-1],
- Tristan P., Abrile P., Massa J., Ferraggine V., Rivero L. y Wainschenker R.(2009).: Evolución en el desarrollo de SIG's hacia herramientas Open Source. ECImag: 2da Escuela y Workshop de Ciencias de las Imágenes.
- Tristan P., Garijo G., Leguizamón G.,Harlouchet L., Gonzalez A. (2018): Una propuesta de ordenamiento vehicular para el transporte de cargas de Puerto Quequén. XVIII Congreso SEPROSUL Semana de la Ingeniería de la Producción Sudamericana. Córdoba, Argentina.. ISSN 2237-3799
- Villar S. A., Torcida S. & Acosta G. G. (2017): Median filtering: a new insight. Journal of Mathematical Imaging and Vision 57(3):1-17 [DOI: 10.1007/s10851-016-0694-0].
- Wang, C. D., & Lai, J. H. (2016). Nonlinear clustering: methods and applications. In Unsupervised Learning Algorithms (pp. 253-302). Springer, Cham.
- Wang, C. D., Lai, J. H., Suen, C. Y., & Zhu, J. Y. (2013). Multi-exemplar affinity propagation. IEEE transactions on pattern analysis and machine intelligence, 35(9), 2223-2237.
- Xie, W., Noble, J. A., & Zisserman, A. (2018). Microscopy cell counting and detection with fully convolutional regression networks. Computer methods in biomechanics and biomedical engineering: Imaging & Visualization, 6(3), 283-292.
- Xu, J., Wang, G., & Deng, W. (2016). DenPEHC: Density peak based efficient hierarchical clustering. Information Sciences, 373, 200-218.
- Zhou, F., De la Torre, F., & Hodgins, J. K. (2013). Hierarchical aligned cluster analysis for temporal clustering of human motion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(3), 582-596.

Facilidades disponibles y/o forma de acceso y fuentes de financiamiento.

El grupo donde se planifica llevar a cabo el proyecto cuenta con 4 oficinas con 10 puestos de trabajo con amoblamiento, equipamiento informático actualizado, impresoras, conectividad a Internet de 1 Gbps, equipamiento para realizar cálculos de alto desempeño y acceso al Centro de Cómputos de Alto Desempeño de la UNICEN. El grupo cuenta con el financiamiento de la UNICEN para este tipo de proyectos. Además, el grupo ha accedido anteriormente a otras fuentes de financiamiento, como proyectos PICT de la Agencia de Promoción Científica y Tecnológica, proyectos de la Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires, proyectos de transferencia con el HIBA para la compra de equipamiento e insumos. Por otro lado, se ha recurrido a financiamiento de proyectos de cooperación Bilateral y las becas BEC.AR del Ministerio de Educación para la realización de estadías de investigación. Para cubrir las becas de grado y postgrado, se recurre a programas de instituciones como el Consejo Interuniversitario Nacional y el CONICET. Se prevé seguir contando con estas fuentes en la medida de lo posible. Además, debido a la reciente incorporación del INTIA como Centro Asociado CICPBA, es posible acceder a líneas de financiamiento específicas. En el marco de la cooperación que ya se tiene con la Universität de Barcelona, se planifica la postulación de estudiantes de postrado a las Becas de la Fundación Carolina.