

"LIDI: Un esquema de las líneas de Investigación y Desarrollo"

Armando De Giusti¹, Patricia Pesado², Marcelo Naiouf³, Laura Lanzarini⁴

Facultad de Informática. Universidad Nacional de La Plata.

Resumen

Se da una visión sintética de las líneas de Investigación y Desarrollo principales del Laboratorio de I/D en Informática (LIDI) de la Facultad de Informática de la UNLP, de modo de contextualizar las diferentes presentaciones realizadas en WICC 2003.

A modo de Introducción se presenta la metodología de Investigación Aplicada que se utiliza y se menciona su aplicación a las actividades de I/D del Laboratorio, así como los convenios y acuerdos que establecen una fuerte vinculación con el medio externo a la Universidad.

Por último se discute la formación de recursos humanos que constituye un objetivo primario de todo grupo de Investigación y Desarrollo inserto en el ámbito Universitario.

Metodología de I/D

El LIDI tiene una metodología de Investigación en Informática Aplicada que puede sintetizarse en los siguientes pasos:

- Identificar un área temática de investigación y desarrollo de interés. Obtener bibliografía e información sobre el estado del tema, tratando de identificar los problemas tecnológicos en el mundo y en el país.

El LIDI trabaja actualmente en una de las áreas de mayor desarrollo en la Informática que es la de Sistemas de Procesamiento Concurrente, Paralelo y Distribuido y sus aplicaciones.

- Obtenida la información de base se realizan Seminarios de discusión y análisis de la clase de problemas correspondiente al área de interés.

Las líneas o subproyectos dentro del área temática mencionada anteriormente se analizan en seminarios del Laboratorio, que incluyen el estudio de las posibles acciones de transferencia de resultados y de vinculación tecnológica con empresas y organizaciones extra-universitarias.

- Identificar una subclase dentro de los problemas a resolver en el área temática y en lo posible caracterizar un problema real cuya solución sea transferible. Encontrar esta aplicación significa ubicar un foco para guiar el desarrollo posterior a resultados concretos.

Dadas las características del área temática general, existen diversas líneas que tienen aplicación directa en el mundo real: tratamiento de señales, reconocimiento de patrones en imágenes y bases de datos, educación no presencial, algoritmos evolutivos, arquitecturas y algoritmos paralelos, análisis de eficiencia en soluciones paralelas y distribuidas, bases de datos distribuidas, sistemas de tiempo real, automatización de oficinas distribuidas, etc.

¹ Profesor Titular Ded. Exclusiva. Investigador Principal CONICET. E-Mail: degiusti@lidi.info.unlp.edu.ar

² Profesor Titular. Prof. Principal CIC. Co-Director de Proyecto. E-Mail: ppesado@lidi.info.unlp.edu.ar

³ Profesor Titular. Co-Director de Proyecto. E-Mail: mnaiouf@lidi.info.unlp.edu.ar

⁴ Profesor Titular. Co-Director de Proyecto. E-Mail: laural@lidi.info.unlp.edu.ar

TE/Fax +(54)(221)422-7707. <http://lidi.info.unlp.edu.ar>

- Investigar las alternativas técnicas de la solución buscando optimizar el empleo del recurso informático y al mismo tiempo interactuar con el mundo real obteniendo una solución específica para el problema concreto. En esta etapa es muy beneficioso el intercambio de información/resultados/líneas de trabajo con otros grupos afines del país y del exterior.
El LIDI tiene actualmente convenios de cooperación científica con Universidades de America Latina, Europa y Estados Unidos, participando de dos proyectos internacionales. Asimismo tiene una decena de convenios con empresas por acciones de transferencia específicas.
- Discutir la generalización y normalización de la solución obtenida. Discutir su eficiencia. Analizar alternativas a profundizar. Contrastar la implementación real de la solución con el usuario o con la especificación de requerimientos establecida en los objetivos originales del proyecto.
Uno de los pasos más importantes en la investigación y desarrollo de una ciencia aplicada como la Informática es la verificación/validación y evaluación de los resultados obtenidos. Claramente la clase de aplicaciones involucrada en el área objeto de investigación no sólo requiere encontrar soluciones “formalmente correctas” sino eficientes en el sentido de ser operativas en arquitecturas reales, dentro de los tiempos de respuesta que exige el mundo real.
- Publicar resultados obtenidos.
En el LIDI la significación de “publicar” excede el concepto de “paper” y se extiende a los registros de propiedad intelectual, las patentes, los Informes Técnicos y el desarrollo mismo de productos concretos relacionados con el área de I/D.

Marco general de los proyectos de I/D en el LIDI

En la actualidad es innegable la importancia y el creciente interés en el procesamiento distribuido y paralelo dentro del espectro de la Ciencia de la Computación. Existen diversas razones para esta realidad, entre las que se pueden citar:

- El crecimiento de la potencia de cómputo, dado en la evolución de la tecnología de los componentes y en las arquitecturas de procesamiento.
- La transformación y creación de algoritmos que explotan al máximo la concurrencia implícita en el problema a resolver, de modo de distribuir el procesamiento minimizando el tiempo total de respuesta. Naturalmente los paradigmas de programación deben adaptarse a los modelos de arquitectura de procesamiento disponible.
- La capacidad del cómputo distribuido/paralelo de reducir el tiempo de procesamiento en problemas de cálculo intensivo o de grandes volúmenes de datos (bases de datos, imágenes, cómputo científico)
- La necesidad de tratar con sistemas de tiempo real distribuidos, donde los requerimientos con relación al tiempo de respuesta son críticos. En particular interesan los sistemas centrados en WEB.
- La existencia de sistemas “inherentemente concurrentes” y/o “inherentemente distribuidos” en los que la adaptación de la solución al mundo real exige un sistema paralelo o distribuido.

Líneas de I/D en desarrollo en el LIDI

A continuación se citan brevemente los tres subproyectos principales del Laboratorio:

SubProyecto P1- Procesamiento Concurrente y Paralelo

Objetivo:

Investigar los problemas de software asociados con la utilización de arquitecturas de procesamiento paralelo, especialmente sistemas multiprocesador distribuidos.

Los temas fundamentales abarcan la especificación, transformación, optimización y verificación de algoritmos concurrentes ejecutables en sistemas paralelos/distribuidos, la optimización de clases de soluciones en función de modelos de arquitectura multiprocesador y las métricas de complejidad y eficiencia relacionadas con el procesamiento paralelo.

Los modelos de arquitectura experimental propuestos (y disponibles para el proyecto) son cuatro: arquitectura multiprocesador homogénea fuertemente acopladas tipo cubo de transputer, arquitectura multiprocesador distribuida (homogénea o heterogénea) débilmente acoplada tipo NOW, multiprocesadores dedicados a tratamiento de imágenes basados en DSPs y arquitecturas multiprocesador de memoria compartida tipo SGI Origin 2000 (Clementina).

Asimismo se ha incorporado una línea de trabajo en simulación y diseño de arquitecturas VLSI orientadas a multiprocesamiento, tendientes a migrar a hardware procesos de administración de recursos compartidos propios de las capas bajas de los sistemas operativos.

Líneas de Estudio/Investigación y Desarrollo:

- Algoritmos paralelos. Paralelización de algoritmos secuenciales. Optimización de algoritmos.
- Lenguajes para procesamiento paralelo.
- Sistemas paralelos. Paradigmas de programación paralela orientados a la arquitectura de soporte.
- Especificación de procesos concurrentes y paralelos. Lenguajes de Especificación.
- Análisis (teórico y práctico) de los problemas de migración y asignación óptima de procesos y datos a procesadores. Migración dinámica. Balance dinámico de carga.
- Métricas del paralelismo. Speedup, eficiencia, rendimiento, isoeficiencia, balance de carga.
- Implementación de soluciones sobre diferentes modelos de arquitectura. Ajuste conceptual del modelo de software al modelo de hardware, de modo de optimizar el sistema paralelo.
- Arquitecturas paralelas virtuales, basadas en redes de procesadores homogéneos y heterogéneos.
- Clusters dedicados y no dedicados. Eficiencia en la solución de aplicaciones mediante clusters.
- Modelos de Memoria en arquitecturas paralelas (compartida, parcialmente compartida, totalmente distribuida).
- Simulación de arquitecturas a nivel circuito integrado. Lenguajes de simulación/especificación.
- Especificación formal de arquitecturas orientadas a multiprocesamiento.

Productos y/o Resultados finales esperados:

- Desarrollar soluciones a problemas de tratamiento masivo de datos (imágenes, video, bases de datos, señales en general) sobre los 4 modelos de arquitecturas multiprocesador expuestos. Evaluar la eficiencia, rendimiento y speedup obtenido con las soluciones propuestas.
- Resolver problemas concretos que requieren procesamiento paralelo en el área médica, en el área de tratamiento de señales en tiempo real, en cómputo numérico y en procesamiento de bases de datos distribuidas. Transferir resultados.
- Plantear modelos de arquitectura VLSI orientados a multiprocesamiento “on-chip” y estudiar su comportamiento teórico mediante simulación. Analizar la migración de algoritmos de software para tratamiento de señales en tiempo real a hardware dedicado.
- Formar recursos humanos en los temas de Procesamiento Concurrente y Paralelo, incluyendo Tesis de Postgrado.

Líneas de I/D en desarrollo en el LIDI

SubProyecto P2- Sistemas de Software Distribuido. Aplicaciones.

Objetivo:

Realizar investigación y desarrollo en temas relacionados con Sistemas de Software Distribuido, particularmente los relacionados con los aspectos de Ingeniería de Software que se orientan al desarrollo e implementación de proyectos de sistemas distribuidos, incluyendo especialmente los temas de métricas y calidad de software.

Estas tareas abarcan aspectos propios de la arquitectura, comunicaciones, estructuración de bases de datos, así como especificación, validación y desarrollo de aplicaciones de software en ambientes distribuidos. Asimismo se considera especialmente la investigación de la Ingeniería de Software de los sistemas distribuidos de tiempo real, considerando las extensiones necesarias en las metodologías y herramientas clásicas con el objetivo de considerar las restricciones de tiempo.

Dos áreas de particular interés son las Bases de Datos Distribuidas donde se investigan especialmente los problemas de replicación y tolerancia a fallas y el área de Tecnología de Software aplicada en Educación, en la cuál se pone énfasis en las aplicaciones de Educación a Distancia y Educación basada en Internet, así como en las métricas aplicadas a calidad en software educativo.

Líneas de Estudio/Investigación y Desarrollo:

- Conceptos de procesamiento distribuido. Arquitectura, comunicaciones y software.
- Lenguajes y ambientes para procesamiento distribuido.
- Ingeniería de Software de sistemas distribuidos.
- Metodologías de especificación, validación y desarrollo de sistemas de software distribuido.
- Extensiones para el tratamiento de sistemas distribuidos de tiempo real.
- Métricas para la evaluación de software. Métricas para la estimación y seguimiento de proyectos.
- Normas de calidad en software. ISO 900X y CMM.
- Herramientas CASE orientadas a sistemas de software distribuido.
- Herramientas de integración y mantenimiento de proyectos distribuidos.
- Reingeniería de sistemas complejos que migran por down-sizing a esquemas cliente-servidor distribuidos.
- Bases de Datos distribuidas. Replicación. Consistencia. Migración de datos. Tolerancia a fallas.
- Tecnología de software aplicada en Educación. Educación a distancia. Educación basada en WEB.
- Métricas de calidad para software educativo.

Productos y/o Resultados finales esperados:

- Desarrollar soluciones a problemas concretos de software de sistemas distribuidos, con énfasis en el desarrollo de metodologías y herramientas específicas para clases de aplicaciones. Evaluar la calidad, eficiencia, y relación de esfuerzo/costo obtenidos con las soluciones propuestas.
- Desarrollo de ambientes y herramientas orientadas a la toma de decisiones en ambientes y organizaciones distribuidas (académicas, empresarias e industriales).
- Desarrollo de herramientas CASE para el desarrollo y mantenimiento de sistemas distribuidos y distribuidos de tiempo real. Desarrollo de herramientas orientadas al mantenimiento y reingeniería de Sistemas, especialmente en ámbitos distribuidos.
- Desarrollo de tecnología para evaluación de calidad de productos de software. Ambientes de aplicación de métricas y normas de calidad en organizaciones de desarrollo de sistemas de soft.
- Implementar soluciones para problemas de Educación a Distancia y Educación basada en la WEB.
- Formar recursos humanos en los temas de Software de Sistemas Distribuidos y Tecnología Informática aplicada en Educación.

Líneas de I/D en desarrollo en el LIDI

SubProyecto P3- Tratamiento de señales e imágenes. Aplicaciones de Tiempo Real

Objetivo:

Realizar investigación y desarrollo en temas relacionados con software para tratamiento de señales, especialmente en problemas de tiempo real.

Estos temas abarcan aspectos clásicos referidos al tratamiento de señales tales como reconocimiento de patrones y análisis de similitud en imágenes, técnicas de clasificación, análisis de cultivos en base a imágenes, tratamiento de video en tiempo real y diagnóstico médico en base a señales de diferente tipo, en los que se utilizan algoritmos basados en métodos clásicos, algoritmos paralelos sobre diferentes modelos de arquitectura multiprocesador y redes neuronales.

En particular interesan las aplicaciones de redes neuronales y redes neuronales evolutivas en problemas de diagnóstico médico basado en imágenes y las técnicas de razonamiento evidencial extendidas para clasificación de cultivos en base a imágenes hiperespectrales y a problemas de robótica industrial.

En todos los casos se investiga la paralelización de soluciones, no sólo para incrementar la eficiencia de los algoritmos, sino para obtener soluciones en tiempo real.

Líneas de Estudio/Investigación y Desarrollo:

- Procesamiento digital de señales.
- Tratamiento de señales en tiempo real.
- Arquitecturas orientadas al tratamiento de señales en tiempo real. En particular DSP.
- Tratamiento de imágenes. Adquisición, Mejoramiento, Segmentación, Clasificación, Reconocimiento de Patrones.
- Análisis de similitud de imágenes.
- Razonamiento evidencial dinámico. Aplicación a clasificación de cultivos.
- Compresión de imágenes y video. Algoritmos para tratamiento de imágenes y video comprimido en tiempo real.
- Redes neuronales. Algoritmos genéticos. Aplicaciones en tratamiento de señales.
- Redes neuronales evolutivas. Aplicaciones a diagnóstico médico. Aplicaciones en data mining, robótica y visión por computadora.
- Ambientes, lenguajes y herramientas de especificación, implementación y verificación de sistemas de tiempo real. Ingeniería de Software de sistemas de Tiempo Real.

Productos y/o Resultados finales esperados:

- Desarrollar soluciones a problemas específicos de tratamiento de señales, en particular en tiempo real. Entre estos problemas, los más significativos se relacionan con clasificación, reconocimiento de patrones y diagnóstico basado en señales de algún tipo.
Evaluar la eficiencia obtenida con las soluciones propuestas, analizando su posible migración a arquitecturas multiprocesador y la transformación de los algoritmos en soluciones paralelas.
- Resolver problemas concretos utilizando redes neuronales, algoritmos genéticos y redes neuronales evolutivas, comparando las soluciones con métodos clásicos. Estudiar la migración de las soluciones mediante redes neuronales sobre arquitecturas multiprocesador.
- Formar recursos humanos en los temas de software para Tratamiento de Señales, Sistemas de tiempo real y Redes Neuronales aplicadas, incluyendo Tesis de Postgrado.

Por último merecen mencionarse como líneas de I/D en desarrollo en el LIDI el proyecto ALFA en Educación a Distancia en conjunto con Universidades de América y Europa, la línea en Tecnología Informática aplicada en Educación que se desarrolla en cooperación con otras Universidades Nacionales y el incipiente desarrollo de una línea en Robótica y Sistemas Industriales de Tiempo Real, que parte de la aplicación de algoritmos paralelos y algoritmos “inteligentes” basados en redes neuronales y redes neuronales evolutivas.

Formación de recursos humanos en el LIDI

Como comentario final, es de hacer notar que actualmente en el LIDI hay 3 Tesistas de Doctorado con fecha de exposición de su Tesis en el 2003 y otros 5 Tesistas de Doctorado con Tesis en desarrollo para los próximos dos años. En todos los casos hay Directores/CoDirectores del exterior.

Asimismo 8 miembros del Laboratorio tienen previsto presentar Tesis de Magíster entre 2003 y 2004 y se desarrollan unas 10 Tesinas de Licenciatura por año, vinculadas con los temas de I/D del Laboratorio.