

Dos experiencias de enseñanza interdisciplinar entre informática y humanidades

Miguel A. Alonso Pardo, Milagros Fernández Gavilanes,
Carlos Gómez Rodríguez, Jorge Graña Gil, Jesús Vilares Ferro
Facultad de Informática, Universidade da Coruña, Campus de Elviña, 15071 A Coruña
Tel: 981 167 000, Fax: 981 167 160, Email: alonso@udc.es

Resumen

Presentamos la experiencia de un grupo de profesores de Informática en la impartición de dos asignaturas marcadamente interdisciplinares que unen dos ámbitos tan diferentes como la informática y las humanidades: una que se imparte en el segundo ciclo de la titulación de Ingeniería Informática, y otra que se imparte en un máster universitario oficial de la Facultad de Filología. Consideramos que la interdisciplinaridad, aunque puede dificultar la implantación de las asignaturas, es también una fuente de oportunidades para los alumnos, que desarrollan así capacidades y competencias para realizar su labor profesional en un mundo en el que las barreras entre disciplinas son cada vez más difusas.

Palabras Clave: Procesamiento del Lenguaje Natural; Recuperación de Información; Informática; Filología.

Abstract

We present the experience of a group of professors of computer science in teaching two markedly interdisciplinary courses that connect two areas as diverse as computer science and the humanities: one that is taught in the second cycle of the Computer Science degree, and another that is taught in an official university master of the Faculty of Philology. We believe that interdisciplinary courses, although it may hinder the implementation of the subjects, are also a source of opportunities for students to develop skills and competencies in order to perform their professional work in a world where barriers between disciplines are increasingly blurred.

Keywords: Natural Language Processing; Information Retrieval; Computer Science; Linguistics.

1. Introducción

La Informática es una disciplina de carácter eminentemente aplicado, y de la que podríamos afirmar que da soporte a las demás ramas del saber, ya que actualmente una gran parte de los avances en el conocimiento requieren de soluciones informáticas, bien para su simulación, bien como herramienta de apoyo a su desarrollo, bien directamente para su implantación. Ello hace que un gran número de ingenieros informáticos, en el ejercicio de su profesión, trabajen en equipos interdisciplinares. Sin embargo, el contacto con otras disciplinas se fomenta poco en el marco de los actuales

planes de estudio de las Ingenierías en Informática, yendo poco más allá de las asignaturas de fundamentos matemáticos o físicos.

En este conexto, en la Sección 2 de este trabajo presentamos la experiencia de un grupo de profesores en la enseñanza de una asignatura de Ingeniería Informática en la cual convergen saberes de Humanidades e Informática. En este contexto, en lugar de centrar la asignatura en los aspectos puramente informáticos, los profesores hemos decidido fomentar la interdisciplinaridad, incentivando en los alumnos el estudio de los aspectos humanísticos de la asignatura, con el propósito de que esto les sirva de entrenamiento para su futuro desarrollo profesional en otros ámbitos donde la interdisciplinaridad sea la norma.

Asimismo en la Sección 3 presentamos una experiencia que se desarrolla en el sentido opuesto: la transmisión de conocimientos informáticos a alumnos del ámbito de las Humanidades en el marco de una asignatura de un máster oficial de la Facultad de Filología. En este caso centraremos nuestras reflexiones en la interacción entre docentes y discentes procedentes de ámbitos tan dispares.

2. Lingüística para informáticos

La primera de las asignaturas analizadas en este trabajo es una asignatura de la titulación de Ingeniería Informática denominada Lenguajes Naturales. Se trata de una asignatura optativa de segundo ciclo que aborda el Procesamiento del Lenguaje Natural (PLN) [8, 14], rama de las ciencias encargada del tratamiento automático del lenguaje humano por parte de la computadora y que, en última instancia, persigue conseguir que la máquina comprenda el lenguaje humano.

2.1. Descripción de la asignatura

La asignatura cuenta con una carga lectiva de 4 créditos ECTS (2.5 teóricos y 1.5 prácticos), y en su diseño se han tenido en cuenta las directrices del Computing Curricula 2001 [1] y su revisión del 2008 [2], que establecen el PLN como una de las unidades (IS7) del área de Sistemas Inteligentes (IS). Su posición en el plan de estudios obedece a su fuerte interrelación con otras asignaturas de la titulación

(Inteligencia Artificial, Compiladores, etc.), permitiendo así al alumno relacionar conceptos, métodos y técnicas impartidos en las mismas. Asimismo, la asignatura persigue combinar tanto aspectos teóricos centrados en el estudio de fenómenos lingüísticos y de los formalismos computacionales para su representación, como aspectos prácticos derivados de la realización de algoritmos y programas de ordenador que transforman dichos fundamentos teóricos en aplicaciones reales. Brevemente, los objetivos generales recogidos en su guía docente son [10]:

- Diferenciar lenguaje natural y artificial.
- Comprender la complejidad del lenguaje humano.
- Familiarizarse con la terminología lingüística.
- Conocer los algoritmos, técnicas y métodos para el tratamiento automático de los fenómenos léxicos, sintácticos y semánticos del lenguaje humano.
- Asumir que el PLN no se puede automatizar completamente, si bien se pueden desarrollar soluciones satisfactorias en la práctica.
- Identificar los aspectos del lenguaje sobre los que incidir para obtener sistemas de PLN útiles.
- Reflexionar sobre los avances y errores en este campo.

En los indicadores de rendimiento del centro, la asignatura presenta una tasa de presentados sobre matriculados del 58,06 % y un 100 % de aprobados sobre presentados, mientras que la media entre las asignaturas optativas de la titulación es del 75,50 % y del 93,60 %, respectivamente.

2.2. Relación con la lingüística

El carácter interdisciplinar de esta asignatura viene dado principalmente por el hecho de que es preciso retomar o introducir conceptos lingüísticos que los alumnos no veían desde secundaria, en particular aquéllos relacionados con las categorías léxicas, el análisis morfológico, la estructura de frase o el análisis sintáctico. A este respecto resulta curioso que si bien los alumnos han estudiado en la titulación las aportaciones de Noam Chomsky a la Teoría de Lenguajes Formales, sin embargo desconocen que es

un lingüista, y que los formalismos que estudiaron para su aplicación en compilación, englobados en la denominada Jerarquía de Chomsky [7], fueron diseñados inicialmente para servir de base formal en la descripción de la sintaxis de los lenguajes naturales. Este caso ejemplifica el hecho de que el PLN es interdisciplinar no sólo porque sea preciso utilizar conocimiento lingüístico para desarrollar soluciones informáticas en este ámbito, sino también porque investigadores del ámbito humanístico, como el propio Chomsky, han realizado contribuciones relevantes a la Teoría de la Computación.

Algo que tratamos de resaltar a lo largo del curso es la tensión existente entre las aportaciones al PLN de los lingüistas y los informáticos, pues mientras los primeros están más interesados en definir una teoría completa del lenguaje, tanto en su vertiente de comprensión (o competencia) como en la de generación (o realización), los segundos están más interesados en la definición e implementación de algoritmos con una complejidad computacional tratable que permitan resolver razonablemente bien problemas prácticos de uso del lenguaje aunque ello suponga utilizar formalismos con una limitada capacidad de descripción de los fenómenos lingüísticos.

A este respecto debemos destacar que mientras que la mayoría de los textos de PLN siguen considerando las gramáticas independiente del contexto como punto de partida inevitable para la realización del análisis sintáctico en sistemas prácticos, los lingüistas han abandonando este formalismo hace décadas. El propio Chomsky abandonó su creación en favor, primero, de la teoría de rección y ligamento [3] y del programa minimalista [4], después. Sin embargo, en los últimos años parece que se está produciendo también un cambio de paradigma en la parte informática del PLN con el auge de los analizadores basados en dependencias [9, 6]. Para explicar los fundamentos de estos analizadores, debemos abandonar las estructuras de constituyentes, que son las que los alumnos han utilizado en su formación preuniversitaria para la realización del análisis sintáctico de oraciones, para pasar a adoptar estructuras lingüísticas basadas en dependencias, como las sugeridas por la Teoría Sentido-Texto [12]. Esto nos permite además mostrar al alumno que tampoco debe descuidar su formación continua en cultura general, pues los conocimientos adquiridos pueden quedar desfasados en relativamente pocos años.

2.3. Relación con la psicología

Es interesante destacar que la interdisciplinariedad del PLN no proviene sólo de la Lingüística, ya que también la Psicología ha hecho una aportación muy relevante al campo, concretamente en el ámbito de de la semántica léxica. Nos estamos refiriendo concretamente a la base de datos léxica WordNet [13], donde las palabras aparecen agrupadas en forma de conjuntos de sinónimos o *synsets*, cada uno representando un concepto diferente. Asimismo entre dichos *synsets* se establecen relaciones léxico-semánticas como la antonimia, hiperonimia, hiponimia, meronimia, etc. Este recurso fue creado en el Laboratorio de Ciencia Cognitiva de la Universidad de Princeton bajo la dirección del psicólogo George Miller con el fin de desarrollar un sistema que fuese consistente con el conocimiento que se había ido adquiriendo sobre la forma en que los humanos procesamos el lenguaje. Sin embargo, donde se ha hecho realmente popular WordNet es en el campo del PLN, en el cual es empleada habitualmente en tareas tales como la desambiguación de los sentidos de las palabras, recuperación de información, clasificación automática, resumen automático, etc.

3. Informática para lingüistas

Tradicionalmente se ha considerado que los estudios filológicos tienen una difícil imbricación en el tejido producido y que la salida natural de los egresados en estas disciplinas era la docencia en los niveles medios del sistema educativo. Sin embargo, lo cierto es que en los últimos años se ha detectado un incremento en la demanda por parte del mercado laboral de personas con una alta capacitación lingüística que sepan actuar en entornos interdisciplinares, posean habilidades comunicativas, dominen las aplicaciones informáticas, sepan gestionar y organizar la información y, en consecuencia, sean capaces de crear contenidos para cualquier medio o soporte.

Con objeto de dar respuesta a esta necesidad, la Facultad de Filología de la Universidade da Coruña implantó recientemente el Máster Universitario en Lengua y Usos Profesionales (LUP) [11], nacido con el fin de formar especialistas en el uso de la lengua como herramienta de comunicación en la sociedad actual. Más concretamente,

en la definición del máster se recoge que sus titulados estarán capacitados para trabajar como gestores de contenidos, documentalistas, redactores, correctores lingüísticos, asesores lingüísticos, mediadores culturales o lingüísticos y como asesores o técnicos culturales. Podrán desarrollar su actividad tanto en empresas (con especial énfasis en editoriales, medios de comunicación, agencias de prensa, gabinetes de comunicación y agencias de publicidad) como en instituciones públicas (con especial énfasis en museos, fundaciones, etc). Llegados a este punto debemos resaltar que puesto que el máster proporciona a los alumnos capacidades y competencias que proyectan la actividad de los egresados más allá de los ámbitos de actuación típicos de los filólogos, mejorando así sus oportunidades dentro del mercado laboral, es importante evitar que la comunidad de graduados en filología sienta que aquéllos que cursan el máster están tratando de salir de la comunidad filológica, ya que esto pondría en marcha efectos perniciosos indeseables [5].

Los contenidos de la titulación se organizan en cinco módulos, cuatro de ellos obligatorios (Informática Aplicada, Gestión del Conocimiento y Recuperación de Información, Recursos Lingüísticos y Nuevas Tecnologías y Comunicación Oral y Escrita 1) y uno optativo (Comunicación Oral y Escrita 2).

Es interesante observar que el máster, al estar dirigido a personas que ya poseen una elevada competencia lingüística, se centra por una parte en aportar a sus alumnos competencias informáticas, y por otra en complementar y adaptar sus competencias lingüísticas previamente adquiridas para adecuarlas al perfil profesional de la titulación. En particular, las competencias informáticas se desarrollan principalmente dentro de los siguientes cursos:

- Digitalización de Información: técnicas y métodos para la creación de representaciones digitales a partir de soportes físicos analógicos.
- Lenguajes de Etiquetado de Documentos: lenguajes de marcado para la estructuración de textos.
- Arquitectura de la Información en Sedes Web: diseño de sitios web y de espacios de información digital.

- Bases de Datos: conceptos básicos de bases de datos.
- Herramientas Informáticas para la Traducción y la Corrección de Textos: utilización de traductores y correctores automáticos, así como la creación, gestión y uso de diccionarios electrónicos.
- Edición Profesional de Textos: uso y manejo de programas de edición profesional.
- Recuperación de Información en Internet: introducción a las tecnologías básicas de Internet y a conceptos básicos de redes de comunicaciones.

En lo que resta de esta sección nos centraremos en la experiencia docente dentro de esta última asignatura, *Recuperación de Información en Internet*, por ser la que imparten los autores, como ejemplo de asignatura interdisciplinar en la que confluyen los intereses y competencias, aparentemente dispares, de profesores (procedentes del ámbito computacional) y alumnos (procedentes del ámbito humanístico).

3.1. Descripción de la asignatura

Recuperación de Información en Internet es una asignatura obligatoria de 3 créditos ECTS enmarcada en el módulo Gestión del Conocimiento y Recuperación de Información. En ella se pretende dotar al alumno de una serie de capacidades, tanto conceptuales como procedimentales, que le permitan conocer qué es una red de comunicaciones, qué es Internet, y cómo se pueden obtener beneficios tangibles de la información que reside en ella mediante el empleo de herramientas y técnicas del ámbito de la Recuperación de Información (RI). Brevemente, los objetivos generales de la asignatura, recogidos en su guía docente [16], son:

- Conocer la estructura interna de una red de comunicaciones, considerando Internet como un caso práctico de especial relevancia.
- Comprender en qué consiste la RI, cuál es su problemática específica, qué tipo de técnicas y modelos se utilizan para resolverla, y cómo se evalúa el rendimiento de un sistema de este tipo.

- Buscar información en el dominio de toda la web, tanto textual como no textual.
- Reformular una búsqueda en la web para mejorar sus resultados.
- Sintetizar los diversos conceptos estudiados en ideas concretas que permitan abordar tareas de perfeccionamiento de un sistema de RI particular, adaptándolo a la problemática específica de un determinado organismo o empresa.
- Considerar posibles técnicas de integración de las características morfológicas y sintácticas de las lenguas, como método de mejora de los buscadores actuales.
- Considerar otro tipo de procesos de nivel superior a la RI, como pueden ser las tareas de extracción de información o búsqueda de respuestas.

La asignatura se comenzó a impartir en el curso 2007/2008, el primer año de implantación del máster y, por tanto, de las materias que lo integran. Aunque por el momento sólo disponemos de los indicadores de rendimiento correspondientes a dicho curso, los resultados han sido positivos: la tasa de presentados sobre matriculados en la asignatura es del 86,96 % mientras que la tasa de aprobados sobre presentados es del 100 %. Ambos indicadores son superiores a la media de la titulación.

3.2. Interacción profesor-alumno

Cuando nos encontramos con una situación en la que el entorno profesional y académico de los profesores (informáticos) y el de los alumnos (filólogos) es tan dispar, es imprescindible establecer desde el primer momento un nexo entre ambos estamentos para lograr llevar a buen puerto el desarrollo de la asignatura. Es de vital importancia evitar la sensación (sea real o subjetiva) de que ambos colectivos están fuera de su ambiente, esto es, de que los profesores dan esa materia por compromiso u obligación y de que los estudiantes se resignan a recibir esos contenidos porque así lo estipula el programa de la titulación. Para ello es preciso dejar claro en la primera sesión presencial que:

1. Los profesores de la asignatura han decidido voluntariamente participar en el máster interdisciplinar.
2. Los profesores consideran importante para su propia formación dar clase fuera de una titulación puramente tecnológica, ya que ello les permite interactuar con usuarios reales de dicha tecnología.
3. Los alumnos no deben sentir inquietud de enfrentarse a una materia tecnológica, aunque su formación sea humanística, y para ello hay que hacerles ver que si se han matriculado en el máster es porque en un determinado momento se han sentido motivados en adquirir tal formación tecnológica y que el profesorado se esforzará por mantener viva esa motivación.
4. El profesorado será permeable a la transmisión de conocimientos por parte del alumnado. Debemos tener en cuenta que en un máster todos los participantes son titulados universitarios y por tanto debería existir un cierto flujo bidireccional de conocimientos entre ambas partes, si bien como en toda acción docente, dicho flujo sea principalmente en sentido profesor-alumno.

En lo que concierne al desarrollo de las sesiones presenciales, muchas de las soluciones propuestas en [15, Sec. 4.3] son aplicables en este caso. En particular, hemos detectado que se puede facilitar la colaboración mediante los mecanismos que se describen a continuación.

En primer lugar, en una asignatura tecnológica como ésta aparecen multitud de términos nuevos, la mayor parte de ellos acrónimos. Realizar una simple enumeración de la forma expandida de tales acrónimos desmotiva al alumnado. Es importante hacerles comprender que lo que se pretende es que incorporen esos términos a su diccionario personal y que por tanto el objetivo es que los manejen con la misma soltura con que manejan las palabras de uso corriente. Para ello se deben generar en clase diálogos en los que dichas palabras ocurran de modo natural. Esta forma de adquirir léxico tecnológico se adapta muy bien a la formación lingüística del alumnado.

En segundo lugar, es necesario resaltar que los sistemas actuales de acceso a la información no son capaces de tratar con toda la complejidad propia del lenguaje

humano [8, 14]. Es importante resaltar los problemas que surgen por el multilingüismo (en qué idioma se realiza una consulta, en qué idioma deberán estar los documentos que queremos recuperar, cómo se puede detectar de manera automática el idioma utilizado, etc.); de las simplificaciones que se hacen a nivel léxico (eliminación de acentos y otras marcas, qué ocurre con las variantes flexivas y derivativas de una palabra, etc.); y del escaso uso que se hace de la información sintáctica (si se tiene o no en cuenta la estructura de la frase, el orden de las palabras, la co-ocurrencia de las mismas, etc.) y semántica (desambiguación de palabras homónimas, etc.). Las discusiones sobre estos aspectos permiten al alumnado darse cuenta de la importancia de la capacitación lingüística que ellos poseen para realizar ciertas tareas documentales, así como de la dificultad que supone proveer de dicha capacitación a un sistema informático y, consecuentemente, de las posibilidades laborales que se abren para ellos.

Finalmente, aunque pueda parecer poco relevante, a la hora de facilitar la interacción con el alumnado ha sido muy importante crear una página web de la asignatura [17], en la cual, más que aspectos administrativos de la misma (datos del profesorado, horarios, bibliografía, etc.) nos hemos centrado en contenidos que contribuyan a los objetivos docentes. En particular, además de mostrar el programa detallado de la asignatura, se proporciona libre acceso a las presentaciones utilizadas en clase y se recomienda la consulta de una serie de sitios web en los que se puede obtener material complementario, entre los que se incluyen, por ejemplo, enlaces a programas de televisión en los que se tratan aspectos sobre los buscadores web, la privacidad de la información, las empresas de Internet, etc. En este punto se ha hecho evidente que, mientras que para el profesorado hacer accesible todo el material docente a través de páginas web es la norma, no ocurría lo mismo con el alumnado, más habituado a un acceso limitado a cierto material fotocopiado proporcionado a posteriori. Paralelamente, esto nos ha permitido identificar dos tipos de comportamiento en el alumnado:

- Por una parte están aquellos estudiantes que han proseguido con su hábitos de trabajo del pasado, consistentes en asistir a la presentación del profesor y

utilizar el material presente en la web para su consulta posterior como material de apoyo en el estudio. Para estos alumnos, la página web no supone más que un cambio de dispositivo de almacenamiento de la información, que pasa a ser un archivo digital en lugar de un conjunto de hojas de papel fotocopiadas, con la importante diferencia de que las fotocopias debían pagarse mientras que el acceso a la web es gratuito. Este hecho puede parecer anecdótico, sin embargo es uno de los mayores incentivos para aceptar la documentación en formato electrónico.

- Por otra parte nos encontramos con aquellos estudiantes que desarrollan un comportamiento proactivo: acceden a la información antes de la clase, la analizan, y la utilizan para plantear dudas durante la presentación por parte del profesor

4. Conclusiones

En el nuevo marco del Espacio Europeo de Educación Superior (EEES) las titulaciones de carácter interdisciplinar serán cada vez más comunes. Para evitar un posible choque de intereses y culturas entre profesorado y alumnado es necesario que ambos estamentos asuman la complementariedad de sus roles, conocimientos y competencias. En este contexto, creemos que las experiencias descritas en este trabajo, que combinan dos ámbitos tan dispares como el tecnológico y el humanístico, puede ser de utilidad en la puesta en marcha y desarrollo de materias en este tipo de titulaciones

5. Referencias

1. *Computing Curricula 2001, Computer Science Final Report*, The Joint Task Force on Computing Curricula, IEEE Computer Society and Association for Computing Machinery (2001).
2. *Computer Science Curriculum 2008: An Interim Revision of CS 2001*, Report from the Interim Review Task Force, IEEE Computer Society and Association for Computing Machinery (2008).

3. N. Chomsky, *Lectures on Government and Binding*, Mouton de Gruyter (1981).
4. N. Chomsky, *El Programa Minimalista*, Alianza Editorial (1999).
5. R. G. Fryer Jr. "A model of social interactions and endogenous poverty traps", *Rationality and Society*, **Vol.** 19 (2007) 335-366.
6. C. Gómez Rodríguez, *Parsing schemata for practical text analysis*. Imperial College Press/World Scientific (2010).
7. J. E. Hopcroft, J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley (1979).
8. D. Jurafsky, J. H. Martin, *Speech and Language Processing. 2nd Edition*. Pearson Education (2009).
9. S. Kübler, R. McDonald, J. Nivre, *Dependency Parsing*, volume 2 of Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers (2009).
10. Lenguajes Naturales, guía docente de la asignatura, https://campusvirtual.udc.es/guiadocente/guia_docent/index.php?centre=614&ensenyament=614111&assignatura=614111625 (visitada el 25 de mayo de 2010).
11. Máster Universitario en Lengua y Usos Profesionales (en gallego). <http://www.udc.es/filo/html/master/master.html> (visitada el 25 de mayo de 2010).
12. I. A. Mel'cuk. "Dependency in natural language". In A. Polgu and I. A. Mel'cuk, editors, *Dependency in Linguistic Description*, chapter 1, pages 1-110. John Benjamins Publishing Company (2009).
13. G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, K. J. Miller, "Introduction to WordNet: An on-line lexical database", *International Journal of Lexicography*, **Vol.** 3 (1990) 235-244.
14. R. Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, Oxford University Press (2005).
15. S. Moore, G. Walsh, A. Risquez, *Teaching at College and University. Effective strategies and key principles*, McGraw-Hill: Open University Press (2007).
16. Recuperación de Información en Internet, guía docente de la Asignatura, https://campusvirtual.udc.es/guiadocente/guia_docent/index.php?centre=613&ensenyament=613421&assignatura=613421107&idioma= (visitada el 25 de mayo de 2010).
17. Recuperación de Información en Internet, página web de la asignatura, <http://www.grupolys.org/docencia/ri/> (visitada el 25 de mayo de 2010).