

imaxin|software

16 anos desenvolvendo aplicações no campo do processamento da linguagem natural multilingue

16 years developing applications for multilingual natural language processing

José Ramom Pichel Campos

imaxin|software

jramompichel@imaxin.com

Diego Vázquez Rey

imaxin|software

diegovazquez@imaxin.com

Antonio Fernández Cabezas

imaxin|software

afernandez@imaxin.com

Luz Castro Pena

imaxin|software

luzcastro@imaxin.com

Resumo

imaxin|software é uma empresa criada em 1997 por quatro titulados em engenharia informática com o objetivo de desenvolver videojogos multimédia educativos e processamento da linguagem natural.

16 anos depois tem desenvolvido recursos, ferramentas, aplicações multilingues para diferentes línguas: Português (Galiza, Portugal, Brasil, etc.), Espanhol (Espanha, Argentina, México, etc.), Inglês, Catalão, Francês.

Neste artigo redigido em português-galego faremos uma descrição daqueles principais fitos deste percurso tecnológico e humano.

Palavras chave

processamento da linguagem natural, correção ortográfica e linguística, correção gramatical e de estilo, dicionários eletrónicos, corpus, tradução automática, opinion mining, biotext mining, sentiment analysis, Microsoft, open source

Abstract

imaxin|software is a company created in 1997 by four computer engineers with the aim of developing educational multimedia games and natural language processing tools.

After 16 years **imaxin|software** has developed resources, tools and applications for different languages, specially for Portuguese (Galiza, Portugal, Brazil, etc.), Spanish (Spain, Argentina, México, etc.), English, Catalan, French.

In this article we will describe the main highlights of this technological and human challenge.

Keywords

natural language processing, spell-checkers, grammar checkers, electronic dictionaries, corpus, machine

translation, opinion mining, biotext mining, sentiment analysis, Microsoft, open source.

1 Introdução

imaxin|software é uma empresa dedicada ao desenvolvimento de serviços e soluções avançadas de software e multimédia desde o ano 1997, especializada em engenharia linguística e videojogos multimédia educativos e formativos.

Dentro da primeira linha de desenvolvimentos, **imaxin|software** é desde o ano 2000 fornecedor de tecnologia linguística para Microsoft. Além disso, podemos destacar entre os principais desenvolvimentos em PLN os sistemas de correção ortográfica, gramatical, estilística; sumários de textos, sistemas de opinion mining, pesquisa semântica, sistemas de codificação médica de histórias clínicas, detecção automática de entidades (NER), bem como a plataforma líder europeia em tradução automática de código aberto: Opentrad¹.

No campo dos multimédia, **imaxin|software** desenvolve desde o ano 1999 sistemas de aprendizagem para meninos e adultos mediante o uso dos videojogos, o que na atualidade é conhecido por serious games. Dentro destes desenvolvimentos multilingues podemos destacar: Vetas (jogo formativo de riscos laborais para as empresas do granito do Porrinho), Climântica (um jogo simulador de cidades sustentáveis), Keco e a Eco-panda (videojogo educativo de educação meio ambiental), minijogos com a plataforma wii e aplicações móveis (Apps).

No campo da localização de software, **ima-**

¹<http://www.opentrad.com>

xin|software encarregou-se de traduções de software especialmente a idiomas como (galego/português, espanhol, inglês, francês) de grande volume como Microsoft Windows XP, Vista ou 7, Office XP ou Office 2013, OpenOffice.org, etc.

2 Linhas de trabalho

imaxin|context

Realização de soluções multilíngues de processamento linguístico e documentário. **imaxin|software** aplica a tecnologia linguística através de ferramentas informáticas na gestão de conteúdos, gestão documentário e gestão do conhecimento, melhorando a produtividade e otimizando a exploração dos recursos intangíveis da empresa:

- **Processamento linguístico:** corretores ortográficos, sintácticos e de estilo, corretores ortográficos em rede, dicionários eletrónicos, tradutores automáticos (OpenTrad).
- **Processamento documentário:** reputação online, pesquisadores documentários, pesquisadores semânticos, sumarizadores de documentos, classificadores automáticos documentários, extractores de informação.

in|gaming

Área multimédia de **imaxin|software** especializada em soluções tecnológicas para a educação, formação e ócio. Contamos com uma equipa multidisciplinar perito na transformação de conteúdos a formato digital e na criação de aplicações inovadoras. Pioneiros no desenvolvimento de jogos educativos multimédia na Galiza, hoje contamos com uma carteira de produtos que englobam serious games, realidade aumentada, aplicações multitácteis, etc.

imaxin|localiza

Criada fruto da experiência na localização de Windows e Office a galego para Microsoft. **imaxin|localiza** presta serviços integrais de localização de software para português europeu (galego, português), espanhol e inglês contando com uma equipa altamente qualificada de linguistas e informáticos. O objetivo deste área é prestar serviço a grandes e medianas empresas de software que precisem traduzir as suas aplicações

para chegar a um maior número de utentes e a mercados internacionais.

3 Certificação CMMI-3

imaxin|software conseguiu a certificação CMMI de nível 3 para duas das três áreas de desenvolvimento de software (Context e ingaming). Resultado desta certificação definiu um conjunto de processos regulares da organização que regem esta atividade e garantem a obtenção de produtos com a qualidade requerida e a satisfação do cliente.

CMMI é um modelo para a melhoria dos processos de uma organização muito complexo de atingir no mundo do software. Aplica-se em áreas de processos que garantem as práticas a seguir pela organização para a consecução dos seus objetivos e se representa por níveis de maturidade atingidos no desenvolvimento da atividade produtiva.

Entre os principais benefícios que proporciona CMMI a **imaxin|software**, estão:

- Alinha o processo Software com a estratégia de negócio da organização.
- Melhora a Qualidade e Produtividade da empresa.
- Antecipa-se os problemas mediante técnicas proativas de gestão.
- Melhora a comunicação na organização, conseguindo uma linguagem comum na mesma.
- Permite que o conhecimento fique na organização.
- Proporciona uns mecanismos de melhoria contínua de processos através de análises de medições.

4 Principais projetos PLN

4.1 Ferramentas de conversão a XML de dicionários e corpus

2013 Projeto: Dicionário da RAG

Este projeto consistiu no desenvolvimento de ferramentas de conversão de um dicionário monolíngue em Word a formato TLEX utilizando a DTD específica para dicionários. Também se desenhou a DTD, realizado a conversão semi-automática de referências cruzadas (Sinónimos, Antónimos, Equivalente). TshwaneLex é uma suíte específica para a elaboração e gestão de dicionários monolíngues, dicionários bilíngues

ou multilíngues em XML. A partir do dicionário convertido a XML-TLEX pode-se optar pela publicação de dicionários em papel, meios eletrônicos ou online. O armazenamento e o manejo de dados lexicográficos utilizando standards da indústria como XML e Unicode, aumenta a produtividade e a qualidade na criação, gestão, revisão e publicação.

4.2 Jogos e Lexicografia

2013 Projeto: Portal das Palavras

O Portal das palavras é um site educativo que põe em valor o dicionário da Real Academia Galega mediante jogos relacionados com as palavras. Com o Portal das Palavras não só melhoraremos a nossa concorrência em idioma galego senão que também aprenderemos jogando. Inclui também o dicionário da RAG com buscas de lemas e sinónimos, vídeos explicativos e guias didáticas para a língua.

4.3 Dicionários monolíngues

1999-2002 Projeto: Dicionário de dicionários contemporâneos

Um dos primeiros Dicionários de dicionários eletrónico da Europa em SGML que permite consultas associadas. Contém 25 dicionários históricos galegos desde o século XVIII ao século XXI, desde Sarmento, até o de Elixio Rivas de 2001 na versão inicial. Sobre este se podem realizar Consultas simples, Consultas complexas, Visualização de entradas, Procuras por refrões, Buscas por poemas, Histórico, Cesta, Hipertexto, Impressão, etc.

Aplicação que trabalha com dicionários previamente convertidos a formato SGML com ferramentas automáticas desenvolvidas adhoc e sobre o que se podem realizar as funções acima indicadas.

2000 Projeto: Dicionário Eletrónico Cumio da Língua Galega

Dicionário eletrónico com perto de 40.000 verbetes. Podem-se realizar as seguintes funções sobre estas: Visualização hipertextual, Seleção de texto, Pesquisas complexas, Impressão, Histórico, Ajuda on-line, Pesquisas por sinónimos, por antónimos etc. Esta aplicação trabalha sobre um dicionário previamente convertido a formato SGML/XML e sobre o que se podem realizar as funções acima indicadas.

2001 Projeto: Dicionário Morris inglês-euskera, euskera-inglês

imaxin|software desenvolveu para o Governo basco o Dicionário de Morris inglês-euskera, euskera-inglês em formato site utilizando diretamente pela primeira vez na Península ibérica a tecnologia XML. Este dicionário eletrónico consta de 40.000 entradas para os dois idiomas presentes no Dicionário regular em Euskadi de inglês-euskera e euskera-inglês. O utente pode fazer consultas por lemas, introduzindo nas caixas de texto correspondentes a palavra, tanto em inglês como em euskera, acedendo à definição do dicionário. Para um determinado lema mostra todos os diferentes verbetes mostrando em ecrã em um formato de saída cómodo para o utente de Internet. Este dicionário site permite a consulta de lemas em inglês e euskera, sendo uma ferramenta fundamental para a internacionalização do euskera.

4.4 Dicionários medievais

2006 Projeto: Dicionário de dicionários medieval

Dicionário eletrónico que contém 13 dicionários construídos a partir de corpus medievais galegos (em estudo da incorporação dos corpus portugueses).

Sobre este se podem realizar Consultas simples, Consultas complexas, Visualização de entradas, Buscas por provérbios, Buscas por poemas, Historial, Cesta, Hipertexto, Impressão, etc. De especial relevância é a construção de tipos de letra próprios medievais.

4.5 Tesouros informatizados contemporâneos

2003 Projeto: Tesouro informatizado da língua galega (Tilga)

Um site que contém textos do galego moderno, desde o ano 1612 à atualidade. Tem 11.409.358 registos e ao redor de 90 mil lemas. Está pensada como corpus de referência onde se podem realizar diferentes consultas: consultas por lema, por palavra, por ano de publicação, por intervalo de anos de publicação, por autor, por obra, etc. É o maior corpus de referência de galego num site e foi realizado para o Instituto da Língua Galega dirigido na altura por D. Antón Santamarina Fernández.

4.6 Tesouros informatizados medievais

2004 Projeto: Tesouro informatizado Medieval da língua galega (TMILG) (Site que contém textos do galego medieval, desde o século XII até o XVII)

Contém um total de 140.000 palavras indexadas em 80 obras, ao redor de 23.000 páginas. Está pensada como um corpus de referência onde se podem realizar diferentes consultas: por palavra, gênero, subgênero, obra específica, intervalo de datas, consultas complexas. As consultas apresentam relatórios estatísticos com informação que abarca desde intervalos de séculos até tipos de documentos nos que aparece em texto consultado. É o maior corpus de referência do galego-português medieval em site.

4.7 Optimizadores semânticos de pesquisas

2008 Projeto: “Optimizador de pesquisas em bibliotecas mediante tesauros”

O objectivo do módulo Optimizador é sugerir sinónimos nas buscas efetuadas polos utentes nos sistemas de consulta bibliográfica do CSBG (Centro Superior Bibliográfico da Galiza). Estes sistemas permitem a consulta de bancos de dados bibliográficas, mostrando resultados sobre os termos procurados. A função do Optimizador é sugerir a possibilidade de alargar essa consulta realizada atendendo a sinónimos do termo consultado. Inicialmente o sistema apresenta os resultados em dois idiomas: galego e castelhano. foi desenvolvido para que em um futuro se possam incorporar outras variantes e idiomas, como o português, inglês, francês etc. A opção de realizar a implementação através de um serviço site é para facilitar a integração do módulo em diferentes tipologias de sites, tanto quanto a linguagens como a plataformas. Está integrado com o software de gestão de bibliotecas de código aberto Koha.

4.8 Hemerotecas e Bibliotecas digitais

2009 Projeto: “Bibliotecas digitais”

O projeto consistiu na digitalização dos boletins históricos da Real Academia Galega, seguindo as adendas A e B das “Diretrizes para projetos de digitalização de coleções e fundos de domínio público, designadamente para aqueles custodiados em bibliotecas e arquivos”, na sua última versão publicada pela Subdirección Geral de Coordinación Bibliotecária do Ministério de Cul-

tura. Ademais atribuíram-se dados e metadatos que codificam as suas descrições e permitem a sua carga em um repositório OAI definido segundo as especificações “The Open Archives Initiative Protocol for Metadata Harvesting” e utilizando a norma ISSO 15836 (Dublin Core). Integração em Galiciana-Hispana-Europeana. Criação de uma aplicação de base de dados em formato site que permita a consulta e visualização dos boletins em formato eletrónico, de forma singela e eficiente.

4.9 Tradução automática

2005-2013 Projeto: “Opentrad: plataforma de serviços de tradução de código aberto”

Clientes: Instituto Cervantes, El Pais, Ministério de Administrações Públicas, Xunta de Galicia, La Voz de Galicia, Faro de Vigo, El Progreso de Lugo, Universidade de Santiago de Compostela, Universidade de Vigo, Universidade da Corunha, CHUAC, Eroski, etc.

Opentrad é a plataforma de tradução automática em código aberto líder no mercado espanhol. Opentrad melhora a comunicação multilíngue, permite publicar informação em diferentes idiomas, reduz custos e os tempos de tradução humana e permite contar com versões multilíngues das aplicações empresariais.

Opentrad está presente em administrações, empresas e portais de Internet traduzindo milhões de palavras diariamente. Ministério de Fazenda e Administrações Públicas, Xunta de Galicia, Instituto Cervantes, La Voz de Galicia, Parlamento da Galiza, NovaGaliciaBanco ou a Kutxa são alguns dos clientes que confiam a sua comunicação multilíngue a Opentrad.

A melhoria contínua do sistema permite-nos oferecer melhor qualidade entre línguas próximas (Espanhol-Francês Espanhol-Português, Espanhol-Português do Brasil, Espanhol-Catalão, Espanhol-Galego, etc.) que outros tradutores automáticos (Google Translate, Systran, Babelfish).

4.10 Serviços

O nosso modelo de negocio está baseado em diferentes serviços ao redor do tradutor automático.

Serviços de tradução automática em aluguer

Opentrad Server (Personalização do Servidor de tradução Opentrad com dicionário personalizado em 150 termos): Ferramentas online: Tradução de textos, Tradução de documentos, Tradução de URL, API de integração.

Opentrad Server Premium

Personalização a medida do cliente (1.000 termos), Ferramentas online: Tradução de textos, Tradução de documentos, Tradução de URL, Ferramentas servidor: API de integração, Pastas dinâmicas, Tradução via e-mail. Integração em Microsoft Word ou Navegadores.

Opentrad na Aplicateca

Desde o passado 2012 está presente na Loja Cloud Aplicateca de Telefónica o que nos permite chegar a todos os clientes de Movistar. Este 2013 está planificado integrar Opentrad em dous novos operadores internacionais.

4.11 Correção ortográfica e gramatical

1998-2012 Projeto: “Galgo e Galgo 2.0”

O corretor imaxin Galgo é o primeiro corretor do mercado que soluciona não só erros de carácter tipográfico e ortográfico, senão também léxicos, que se podem encontrar em textos escritos em galego. Esta nova versão de Imaxin Galgo, aplicação pioneira para a correção de textos em galego, atualiza e melhora esta ferramenta pensada para a análise dos problemas de carácter ortográfico e léxico intrínsecos a todo texto escrito no nosso idioma.

Assim, é possível que se misturem num mesmo documento desde castelhanismos, verbos mau conjugados, vulgarismos, erros propriamente ortográficos ou até intersecções destas tipologias em uma mesma palavra. Entre as atualizações, destaca a modificação da base de dados segundo a norma aprovada no ano 2003 pola Real Academia Galega e o Instituto da Língua Galega.

4.12 Correção de linguagem não sexista

2008 Projeto: “Exeria: corretor de linguagem não sexista” [em parceria com Tagen Ata]

O corretor de linguagem não sexista é uma ferramenta informática integrável no pacote ofimático livre OpenOffice que tem como finalidade ajudar a realizar documentos com uma linguagem não-sexista. Exeria oferece uma ajuda interativa para a edição de textos que reflitam um tratamento igualitário da linguagem. Exeria nasceu dentro de um plano do Governo galego para a igualdade entre mulheres e homens 2007-2010. Exeria pretende ajudar a construir uma linguagem que visibilize a presença das mulheres no colectivo e integrar a sua consideração enquanto agentes par-

ticipativos nos diferentes âmbitos sociais. Com esta finalidade, Exeria facilita a construção de um discurso mais inclusivo oferecendo alternativas e soluções para aqueles termos que podem, em alguns contextos, implicar um uso discriminatório no que atinge à linguagem.

5 Principais projetos I+D

CELTIC (Conhecimento estratégico liderado por tecnologias para a Inteligência Competitiva)

[Em desenvolvimento]: O projeto está orientado no campo da vigilância tecnológica e o Social Média Marketing. Participantes no projeto: IN-DRA, Elogia, imaxin|software, Saec-data, USC-GE, USC-CA, Gradiant. Financiado polo programa FEDER-INNTERCONNECTA, através do CDTI.

Coruxa Biomedical Text Mining: Extrator e codificador automática de informação médica relevante mediante uso-o da engenharia linguística em código aberto

Direção-geral de I+D+i. Xunta de Galicia. Pesquisador principal: imaxin|software, USC-GE, IXA Taldea, Doutor Qsolutions. Transferência ao sector produtivo: talento de codificação SNOMED-CT para histórias clínicas. Financiado polo Programa I+D+i Galego.

Coati Opinion mining: Pesquisa avançada multilíngue em blogues para a recuperação de opiniões e tendências para ou âmbito empresarial e dá administração pública. Direção Xeral de I+D+i. Xunta de Galicia. Pesquisador principal: imaxin|software, USC-GSI, USC-GE, UDC-IR-Lab. Transferência ao sector produtivo: desenvolvimento de ETLs, talento de recuperação de opiniões em blogues para a administração pública e o sector empresarial. Financiado polo Programa I+D+i Galego.

EurOpenTrad “Traducción automática avanzada de código abierto para la integración europea de las lenguas del Estado español”

(PROFIT-350401-2006-5), 2006-2007-2008. Investigador principal: imaxin|software, Eleka, Elhuyar, IXA Taldea, TALP (UPC), Transducens- UA, SLI. Transferência ao sector produtivo: adaptação do tradutor automático OpenTrad a UNICODE, incorporação de detectores de idiomas, detectores de nomes próprios automático. Financiado polo Programa AVANZA-PROFIT do Ministério de Indústria.

OpenTrad “Traducción Automática de Código Abierto para las Lenguas del Estado Español”

(PROFIT-340101-2004-0003, PROFIT-340001-2005-2), 2005-2006. Investigador principal: Eleka, Elhuyar, IXA Taldea, TALP (UPC), Transducens-UA, **imaxin**|software, SLI-Universidade de Vigo. Transferência ao sector produtivo: tradutor automático de espanhol-catalão, espanhol-galego e espanhol-euskera desenvolvido em código aberto com transferência de recursos linguísticos galego e português e motor de tradução adaptado a estes dois idiomas. Financiado polo Programa AVANZA-PROFIT do Ministério de Indústria.

EixOpenTrad “Tradução automática avançada de código aberto entre as variantes do português de Portugal e do português da Galiza”

2006-2007. Investigador principal: **imaxin**|software, Universidad de Santiago de Compostela y SLI. Transferência ao sector produtivo: tradutor automático de galego-português e português-galego desenvolvido em código aberto com transferência de recursos linguísticos galego e português e motor de tradução adaptado a estes dois idiomas.

RecursOpenTrad “Recursopentrad: recursos lingüístico-computacionais de tradución automática avanzada em código aberto para a integración europea da lingua galega”

Pesquisador principal: **imaxin**|software, TALP (UPC), Transducens-UA, SLI. Transferência ao sector produtivo: melhoria do tradutor inglês-galego RBMT, criação de um protótipo estatístico SMT de tradução automática inglês-galego/português, desenvolvimento de ETLs, integração de detectores de idiomas, detectores de entidades no motor apertium.

Extração da informação: “Estudio de necesidades e xerazón de recursos e ferramentas intelixentes em xestión da información e enxeñaría lingüística para a mellora das empresas exportadoras galegas”

Subvenciona: Xunta de Galicia, ref. PGIDT03TICC22Y. Entidades participantes: **imaxin**|software, CESGA (Centro de Supercomputación da Galiza), SLI, Cidadanía rede de aplicacións sociais. Pesquisador responsável: José Ramom Pichel (**imaxin**|software), Xavier Gómez Guinovart (SLI-Uvigo). Transferência ao sector produtivo: realização de um protótipo de

recuperador de informação a partir de um site crawler libertado com licença GPL.

Etiquetagem de textos e desambiguação automática: “Estudo e adquisición de recursos básicos de lingüística computacional do galego para a elaboración e mellora de aplicacións informáticas de tecnoloxía lingüística”

Subvenciona: Secretaria Xeral de Investigación e Desenvolvemento, Xunta de Galicia, 2001-2004 (ref. PGIDT01TICC06E). Pesquisadores principais: José Ramom Pichel (**imaxin**|software), Xavier Gómez Guinovart (SLI-Uvigo). Equipa de projeto: Elena Sacau (SLI), Ángel López (**imaxin**|software). Transferência ao sector produtivo: o lexicón gerado neste projeto foi reutilizado posteriormente para a realización do corretor ortográfico de OpenOffice.org que foi liberto com licença GPL.

“Estudo do erro gramatical para o galego”

Subvenciona: **imaxin**|software, Projeto de I+D (Universidad - Empresa), 2002-2003. Investigador principal: Xavier Gómez Guinovart (SLI-Uvigo). Transferência ao sector produtivo: este projeto inicial foi necessário para desenvolver entre o ano 2006 e 2007 o corretor gramatical Golfinho.

6 Projetos de I+D+i em desenvolvimento

CELTIC: Conocimiento Estratégico Liderado por Tecnologías para la Inteligencia Competitiva (FEDER-INNTERCONECTA)

Os FEDER INNTERCONECTA são projetos Integrados de desenvolvimento experimental altamente competitivos, com carácter estratégico, de grande dimensão e que tenham como objectivo o desenvolvimento de tecnologias novas em áreas tecnológicas de futuro com projeção económica e comercial a nível internacional, supondo ao mesmo tempo um avanço tecnológico e industrial relevante para as autonomias destinatárias das ajudas do “Programa Operativo de I+D+i por e para o benefício das empresas - Fundo Tecnológico,” como é o caso da Galiza.

imaxin|software conseguiu no ano 2012 um projeto FEDER-INNTERCONECTA com um consórcio formado polas seguintes empresas e Universidades: Indra, Elogia, SaecData, Gradient, USC-PRONAT-L (USC), ComputationalArchitecture Group (USC).

Objectivo principal do projeto

Desenvolvimento de tecnologias capacitadoras que facilitem ao tecido empresarial a tomada de decisões estratégicas em tempo quase-real, a partir do conhecimento tanto do meio científico-tecnológico como dos impactos económicos presentes e futuros. Ou o que é o mesmo, o desenvolvimento de tecnologias capacitadoras para a Inteligência Competitiva nas organizações.

As tecnologias a desenvolver durante o projeto cobrirão o processo completo da Inteligência Competitiva, nas suas respectivas fases: agregação de informação, análise da informação extraíndo dela o conhecimento necessário, e a distribuição mediante mecanismos de visualização e iteração avançados para facilitar a tomada de decisões estratégicas.

Aplicações do projeto

Marketing: A competitividade atual imprime a necessidade de dispor de sistemas de monitorização inteligente e em tempo real de redes sociais e análises do impacto dos produtos de uma marca determinada no consumidor, mediante tecnologias avançadas de processamento da linguagem natural e tecnologias semânticas.

Vigilância tecnológica: os desenvolvimentos a realizar neste projeto permitirão o acesso e gestão em tempo real dos conhecimentos científicos e técnicos às empresas, bem como a informação mais relevante sobre o seu contexto, junto ao entendimento a tempo do significado e envolvimento das mudanças e novidades no meio. Isto é indispensável na tomada de decisões das empresas para o desenvolvimento de um novo produto, serviço ou processo para uma organização.

7 Prémios

- Prémios Eganet 2006: Prémio especial a 10 anos.
- Prémios AETIC 2007: Galardoado Opentrad como a melhor aplicação TIC do 2007.
- Prémios Eganet 2008: Prémio à melhor iniciativa de Comunicação Site Institucional polo jogo Keco e procura da Ecopanda
- Prémios Eganet 2008: Finalista na categoria Software Livre pola aplicação Opentrad.
- Prémios Eganet 2009: Prémio Melhor ambiente Laboral
- Prémios Eganet 2009: Finalista na categoria Cultura Digital polo jogo “Climántica”
- Prémio Eganet 2010: Prémio Educação Digital polo projeto “O Valor de IGU” pro-

grama interativo de educação em valores através de videojogos.

- Prémio Leixa-prem 2012 polo uso habitual do galego na empresa
- Prémio Ada Lovelace a Luz Castro Pena polo Colexio de Enxeñeiros/as en informática.

Agradecimentos

A todas as pessoas que têm partilhado connosco este caminho de aprendizagem humana e tecnológica.

Bibliografía

- Aguirre Moreno, José Luis, Alberto Álvarez LUGRÍS, Luz Castro Pena, Xavier Gómez Guinovart, Angel López López, José Ramom Pichel Campos, Elena Sacau Fontenla, e Lara Santos Suárez. 2003a. Adquisición de recursos básicos de lingüística computacional del gallego para aplicaciones informáticas de tecnología lingüística. *Procesamiento del Lenguaje Natural*, 31:303–304.
- Aguirre Moreno, José Luis, Alberto Álvarez LUGRÍS, Iago Bragado Trigo, Luz Castro Pena, Xavier Gómez Guinovart, Santiago González Lopo, Angel López López, José Ramom Pichel Campos, Elena Sacau Fontenla, e Lara Santos Suárez. 2003b. Alinhamento e etiquetagem de corpora paralelos no CLUVI (Corpus Lingüístico da Universidade de Vigo). Em José João Almeida, editor, *CP3A 2003, Corpora Paralelos: Aplicações e Algoritmos Associados*, pp. 33–47, Universidade do Minho, Braga, Portugal.
- Alegría Loinaz, Iñaki, Iñaki Arantzabal, Mikel L. Forcada, Xavier Gómez Guinovart, Lluís Padró, José Ramom Pichel Campos, e Josu Waliño. 2006. OpenTrad: Traducción automática de código abierto para las lenguas del estado español. *Procesamiento del Lenguaje Natural*, 37:357–358.
- de Moura Barros, António Carlos, Angel López López, e José Ramom Pichel Campos. 2008. TMILG: tesouro medieval informatizado da lingua galega. *Procesamiento del Lenguaje Natural*, 41:303–304.
- Gamallo, Pablo e Jose Ramom Pichel. 2008. Learning Spanish-Galician translation equivalents using a comparable corpus and a bilingual dictionary. *Lecture Notes in Computer Science*, 4919:423–433.

- Gamallo, Pablo e José Ramom Pichel. 2005. An approach to acquire word translations from non-parallel text. *Progress in Artificial Intelligence, LNAI*, 3808.
- Gamallo, Pablo e José Ramom Pichel. 2007. Un método de extracción de equivalentes de traducción a partir de un corpus comparable castellano-gallego. *Procesamiento del Lenguaje Natural*, 39:241–248.
- Gamallo, Pablo e José Ramom Pichel. 2010. Automatic generation of bilingual dictionaries using intermediary languages and comparable corpora. Em *Cycling2010*.
- Gamallo Otero, Pablo, Marcos Garcia, e José Ramom Pichel Campos. 2013. A method to lexical normalisation of tweets. Em *Tweet Normalization Workshop at SEPLN*.
- Malvar, Paulo e José Ramom Pichel Campos. 2010. Obtaining computational resources for languages with scarce resources from closely related computationally-developed languages. the Galician and Portuguese case. Em *II Congreso Internacional de Lingüística de Corpus (CILC10)*, Universidade da Coruña.
- Malvar, Paulo e José Ramom Pichel Campos. 2011a. Generación semiautomática de recursos de opinion mining para el gallego a partir del portugués y el español. Em *ICL11: Workshop on Iberian Cross-Language NLP Tasks*.
- Malvar, Paulo e José Ramom Pichel Campos. 2011b. Métodos semiautomáticos de generación de recursos de opinion mining para el gallego a partir del portugués y el español. *Novática: Revista de la Asociación de Técnicos de Informática*.
- Malvar, Paulo, José Ramom Pichel, Óscar Senra, Pablo Gamallo, e Alberto García. 2010. Vencendo a escassez de recursos computacionais. Carvalho: Tradutor automático estatístico Inglês-Galego a partir do corpus paralelo Europarl Inglês-Português. *Linguamática*, 2(2):31–38.
- Pichel, José Ramom, Paulo Malvar López, Oscar Senra Gómez, Pablo Gamallo Otero, e Alberto García. 2009. Carvalho: English-Galician smt system from english-Portuguese parallel corpus. *Procesamiento del Lenguaje Natural*, 41.
- Pichel Campos, José Ramom. 1996. Problemas e solucións para a terminoloxía galega. Em *I Congreso Internacional da lingua galega (ed. ILG)*.
- Pichel Campos, José Ramom. 1997. Funciones terminológicas en la lengua gallega: problemas metodológicos y soluciones al respecto. *Uzei, Centro Vasco de Terminología y Lexicografía*.
- Pichel Campos, José Ramom e Antonio Fernández Cabezas. 1999. “imaxin Galgo v.1.0. *Procesamiento del Lenguaje Natural*, 25:225–226.
- Pichel Campos, José Ramom e Antonio Fernández Cabezas. 2002. Dicionario de dicionarios da lingua galega. *Procesamiento del Lenguaje Natural*, 26:99–100.