

# TICAI 2010



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

TIN2009-07333-E/TSI



**IEEE**

Sociedad Educación  
CAPITULO ESPAÑOL

CD-R 52X



**IEEE** Education  
Society Chapter  
PORTUGAL SECTION

**TICAI 2010 (TICs Aplicadas para el Aprendizaje de la Ingeniería)**

Editores: Manuel Caeiro Rodríguez, Ricardo Azambuja Silveira,  
Paula Escudeiro

Diseñador: Diego Estévez González

ISBN: 978-84-8158-548-3

<http://webs.uvigo.es/cese/TICAI>



# Índice

<b>1. Inovação curricular com ferramentas Web 2.0: uma experiência de Webradio educativa</b>	<b>1</b>
Vítor Manuel Santos Diegues, Clara Pereira Coutinho	
<b>2. A-TVDBR: Um Modelo de Atividades de Aprendizagem no Contexto de Educação a Distância para a TV Digital Brasileira</b>	<b>7</b>
Mozart de Melo Alves Júnior, Arturo Hernández Domínguez	
<b>3. Herramienta para la generación de escenarios de apoyo a la docencia de la seguridad en redes</b>	<b>15</b>
Enrique de la Hoz, Ivan Marsa-Maestre, Miguel A. López-Carmona, María Teresa López-Merayo	
<b>4. Tecnologias da Informação e Comunicação no apoio à aprendizagem no Ensino Superior</b>	<b>25</b>
Bertil P. Marques, Jaime E. Villate, Carlos V. Carvalho	
<b>5. Uso de juegos digitales educativos como herramienta de soporte para el aprendizaje de algoritmos</b>	<b>33</b>
Julián Moreno	
<b>6. Prácticas de control digital de convertidores de potencia emulados con amplificadores operacionales</b>	<b>39</b>
Jacinto M. Jiménez, Miguel, Moreno, Esther de Jódar, Jose A. Villarejo	
<b>7. O Quadro Interactivo Como uma Ferramenta Colaborativa no Contexto do Ensino Português</b>	<b>47</b>
Ana Pinto, Eduardo Cardoso, Paula Escudeiro	
<b>8. Adaptação de Metadados para Repositórios de Objetos de Aprendizagem</b>	<b>55</b>
Liane M. R. Tarouco, Marcelo A. R. Schmitt	
<b>9. Levitador magnético de muy bajo coste con fines didácticos</b>	<b>61</b>
José I. Artigas, Luís A. Barragán, Sergio Llorente, Álvaro Marco	
<b>10. Personal Learning Environments no contexto virtual de um Mestrado em Comunicação Educacional Multimédia</b>	<b>67</b>
José Bidarra, Ana Margarida Sousa, Filomena Grazina, Paula Simões, Paulo Azevedo	
<b>11. Uma Abordagem Computacional para Construção de Mapas Conceituais a partir de Textos em Língua Portuguesa do Brasil</b>	<b>75</b>
Juliana Hiroko Kowata, Davidson Cury, and Maria Claudia Silva Boeres	
<b>12. Esquemas de enseñanza-aprendizaje del concepto de la comunicación solapada de mensajes en redes de comunicación</b>	<b>81</b>
Elsa Macías, Álvaro Suárez	
<b>13. Um Objeto de Aprendizagem para apoio a Aprendizagem de Física</b>	<b>91</b>
Maria de Fátima Webber do Prado Lima, Douglas Hecher and Elisa Boff	
<b>14. Image Easy Sonar: Software para el Procesado Digital de Imágenes Sónar con Aplicaciones Docentes</b>	<b>101</b>
Isabel Zamanillo Sainz de la Maza, José M <sup>a</sup> Zamanillo, Elías Revestido, Francisco Jesús Velasco	
<b>15. Los Recorridos de Estudio e Investigación como productos de Ingeniería didáctica</b>	<b>107</b>
C. Fonseca, A. Pereira, J. M. Casas	
<b>16. Geometria através de Hipertextos com Animações Interativas</b>	<b>117</b>
Maria Alice Gravina e Marina Menna Barreto	
<b>17. Uso de la Pizarra Digital y los Sistemas de Respuestas para la Implantación del Aprendizaje Basado en Proyectos en Ingeniería Informática</b>	<b>123</b>
José Santa, Antonio J. Jara, Germán Villalba, Miguel A. Zamora	
<b>18. Sentiment Analysis on a corpus of texts written by primary school children</b>	<b>129</b>
Eladio Blanco, Fernando Martínez, Antonio Pantoja, Alfonso Ureña	
<b>19. Correção semi-automática de respostas em texto livre</b>	<b>137</b>
Nuno Escudeiro, Paula Escudeiro	
<b>20. Mobile Learning Engine Moodle adaptado aos diferentes Estilos Cognitivos utilizando Hipermedia Adaptativa</b>	<b>143</b>
Patricia Mariotto Mozzaquatro, Roseclea Duarte Medina	

- 21. Tengo una respuesta para usted sobre estilos de aprendizaje, creencias y cambios en los estudiantes** 151  
M. Asunción Castaño, Mercedes Marqués, Rosana Satorre, Antoni Jaume i Capó, David López
- 22. Comunidades de prática em ambientes virtuais: da teoria à experiência colaborativa** 159  
Soraia Maria João Macário, Ticiania Tréz, Sandra Ferrão-Lopes, Jorge Gonçalves, Isabel Cabrita, Lúcia Pombo
- 23. Plataforma Educacional con Remolcador de Altura Autónomo** 167  
F. J. Velasco, E. Revestido, E. López, E. Moyano, A. Bascónes, I. Zamanillo, J. M. Zamanillo
- 24. Recursos Pedagógicos para Dispositivos Móveis: uma Análise com Foco na Matemática** 173  
Silvia Cristina F. Batista, Patricia Alejandra Behar, Liliana Maria Passerino
- 25. Mejorando el aprendizaje de Procesadores de Lenguaje mediante visualizaciones** 181  
Francisco J. Almeida Martínez, Jaime Urquiza Fuentes, J. Ángel Velázquez Iturbide
- 26. FindYourHelp: um módulo de busca por especialistas no AVA Moodle** 189  
Marcos L. Santos, Laís do N. Salvador, Daniela S. Cruzes
- 27. ¿Investigamos correctamente sobre la baja matriculación de mujeres en Informática?** 201  
Agustín Cernuda del Río, Miguel Riesco Albizu
- 28. Implementación de materias comunes en el plan de estudios de grado en Ingeniería Electrónica y Automática de la Universidad de Zaragoza** 209  
C. Bernal Ruiz, A. Bono Nuez, J.M. López Pérez, A. Otín Acín, F.J. Pérez Cebolla, B.Martín del Brío, T. Pollán Santamaría

# Capítulo 18

## Sentiment Analysis on a corpus of texts written by Primary School children

Eladio Blanco, Fernando Martínez, Antonio Pantoja, L. Alfonso Ureña

**Abstract—** Here is described the making of the e-Culturas corpus from written texts by 10-11-year-old-children. These texts have been written from the memory of some event in which the kid has felt fear, happiness or sadness. The purpose of the corpus in this report is the training of an automatic classifier for these sentiments. The results of the evaluation of the classifier are hopeful, in spite of being in a preliminary development status. We think that they show that the methodology used for the making of the corpus is suitable, and specific techniques can perform at a level good enough to be useful within the e-Culturas web platform.

**Index Terms—** Automatic classification, making of corpus, opinion mining, sentiment analysis.

**Abstract—** En este artículo se describen las actividades que se han realizado para generar un corpus de textos acerca de las emociones alegría, miedo y tristeza por niños de entre 10 y 11 años participantes en la edición del 2009 de la Red Internacional e-Culturas. También se detallan los pasos llevados a cabo para refinar los textos que forman el corpus, así como una experimentación preliminar con el corpus y las ideas para un trabajo futuro.

**Index Terms—** Clasificación automática, creación de corpus, minería de opinión, análisis de sentimientos.

Este trabajo fue presentado originalmente al CONGRESO “5ª Conferencia Ibérica de Sistemas y Tecnologías de Información”, Santiago de Compostela (España).

Manuscript received April 14, 2011. (Write the date on which you submitted your paper for review.) This work was supported in part by the U.S. Department of Commerce under Grant BS123456 (sponsor and financial support acknowledgment goes here). Paper titles should be written in uppercase and lowercase letters, not all uppercase. Avoid writing long formulas with subscripts in the title; short formulas that identify the elements are fine (e.g., “Nd-Fe-B”). Do not write “(Invited)” in the title. Full names of authors are preferred in the author field, but are not required. Put a space between authors’ initials.

E. Blanco is with the IDEO Group at University of Jaen, Spain (e-mail: elbloo@gmail.com).

F. Martínez and L. A. Ureña are with the SINAI Group at University of Jaen, Spain (e-mail: dofer@ujaen.es / laurena@ujaen.es).

A. Pantoja is with the IDEO Group at University of Jaen, Spain (e-mail: apantoja@ujaen.es).

### I. INTRODUCCIÓN

El análisis de sentimientos (*sentiment analysis*) es una tarea del Procesamiento del Lenguaje Natural que se preocupa por el tratamiento automático de opiniones, sentimientos y subjetividad presente en el texto escrito, [1] (Pang & Lee, 2008). La popularización de lo que ha venido a denominarse Web 2.0 requiere de herramientas adecuadas que permitan explotar de una manera adecuada tal flujo de información. Es prometedora la confluencia de toda esta nueva fuente de texto subjetivo junto con herramientas que permitan modelar automáticamente el conocimiento allí expresado de tal forma que sea computacionalmente manipulable.

Por Educación Emocional se entiende el desarrollo de conocimientos y habilidades sobre las emociones con la finalidad de capacitar a la persona para afrontar mejor los retos de la vida, mejorando su bienestar personal y social, [2] (Bisquerra, 2002).

En el presente trabajo se propone la aplicación de técnicas propias de análisis de sentimientos como una herramienta valiosa para la integración intercultural y desarrollo emocional de niños de 10 y 11 años dentro de la Red Internacional e-Culturas, [3] (Pantoja, 2008) y [4] (Alcaide, Blanco, Pantoja, & Jiménez, 2008). Este trabajo, es parte de uno de mayor envergadura, cuya finalidad será la detección automática de emociones en el texto que escriban los alumnos de la Red Internacional e-Culturas acerca de una cuestión intercultural o sobre Educación Emocional para, dependiendo de la emoción detectada, plantearle al alumno unas actividades que intenten corregirla si esta ha resultado negativa.

El artículo está organizado como sigue: a continuación, una pequeña descripción de la Red Internacional e-Culturas. En la sección III se justifica la elección de las emociones a detectar, en la sección IV se presentan las actividades utilizadas para generar el texto del corpus, en la sección V se detallan los pasos para la creación del corpus, en la sección VI se presenta una breve experimentación con el corpus.

TABLA I  
NÚMERO DE ALUMNOS POR PAÍS EN E-CULTURAS 2009

País	Número de alumnos
España	139
Argentina	55
Brasil	121
Ecuador	157
Paraguay	48
<b>Total</b>	<b>517</b>

Finalmente en la sección VII se resume brevemente este trabajo, así como la línea de investigación futura.

## II. RED INTERNACIONAL E-CULTURAS

El proyecto e-Culturas [5] desarrolla desde hace años el uso de las TIC en la orientación y la tutoría, y la educación intercultural.

El objetivo principal es favorecer la interacción y el conocimiento cultural entre el alumnado y profesorado de centros educativos de España y de Latinoamérica. Éste se completa con varios objetivos específicos, entre los que destacan:

- Crear una red intercultural entre los países participantes (España, Argentina, Brasil, Chile y Paraguay).
- Conocer los efectos del Programa Intercultural e-Culturas en el cambio de las actitudes y las competencias comunicativas interculturales del alumnado de los centros participantes.
- Favorecer la interacción y el conocimiento cultural entre alumnado y profesorado de centros educativos españoles y latinoamericanos de nivel primario, mediante el diseño y aplicación de un programa educativo que utilice como soporte las TICs.

Los sujetos que han participado en el estudio fueron 517 de 5 países diferentes agrupados en 14 centros y 18 aulas (cada grupo contaba con un tutor que les ayudaba en caso de alguna dificultad y a la vez los tutores de los grupos hermanados estaban también en contacto para cualquier problema que les surgiera). Todos eran alumnos de 6º de Educación Primaria en España y de niveles similares en el resto de los países (10 y 11 años de edad). La Tabla I muestra los alumnos que han participado en esta experiencia en la última edición de e-Culturas, desglosado por países.

## III. TRABAJO RELACIONADO

La llegada de la Web 2.0 [10] (O'Reilly, 2005) ha multiplicado el contenido textual en Internet, pues todos sus usuarios toman parte activa como productores de información. Pero esto trae consigo la consecuente diversificación de los registros lingüísticos lejos de formalismo o los lenguajes académicos. Comentarios en "blogs", opiniones en tiendas "online", discusiones en foros y otros paradigmas de publicación han atraído el interés por

TABLA II  
CLASIFICACIÓN DE LAS EMOCIONES MÁS EXTENDIDAS

Emoción	Goleman (2001)	Bisquerra (2002)
Ira	X	X
Tristeza	X	X
Miedo	X	X
Placer	X	
Amor	X	X
Sorpresa	X	X
Enojo	X	
Vergüenza	X	X
Ansiedad		X
Aversión		X
Alegría		X
Humor		X
Felicidad		X
Esperanza		X
Compasión		X
Cólera		X

hacer minería y análisis sobre estos textos de libre escritura y, en consecuencia, retador análisis [11] (Quarterly, 2009).

Esta revisión de la minería de textos se conoce como Minería de Opiniones y Análisis de Sentimientos (Opinion Mining and Sentiment Analysis) [1] (Pang & Lee, 2008). Si bien ambos términos se usan de forma no diferenciada, no son lo mismo. El análisis de sentimientos trabaja sobre aspectos como la clasificación de la polaridad de los sentimientos o la subjetividad expresados en un texto. Evidentemente, estos análisis son de gran utilidad para las tareas de minería de opinión, en las que se intenta obtener, a partir de los comentarios escritos por diversos usuarios, una idea general de la opinión del público sobre productos, noticias u otras entidades presentes en la red.

El interés que despiertan estas tareas puede verse reflejado en la aparición de recursos, como una variante de WordNet que refleja aspectos relativos a sentimientos en las palabras [12] (Esuli & Sebastiani, 2006) o corpus destinados a ser usados como base para la experimentación en técnicas de minería de opiniones y análisis de sentimientos, como el corpus multilíngüe de NTCIR [13] (Seki, Evans, Ku, Chen, Kando, & Lin, 2007) o el corpus usado en las competiciones TREC [14] (Ounis, MacDonald, de Rijke, Mishne, & Soboroff, 2006). Además, ya han tenido lugar importantes competiciones a nivel internacional sobre la materia, como las cuatro últimas ediciones del TREC Blog Track (de 2006 a 2009) [15] (Ounis, Macdonald, & Soboroff, 2008) o el MOAT de 2008 [16] (Passant & Laublet, 2008).

## IV. ELECCIÓN DE LAS EMOCIONES

En la literatura existen diversas clasificaciones de emociones, [2] (Bisquerra, 2002) y [6] (Goleman, 2001). En la Tabla II se muestran las clasificaciones de sentimientos

más extendidas. En este trabajo se han seleccionado tres de ellas atendiendo, primero, a su utilidad de cara a su integración dentro de la plataforma de e-Culturas y, segundo, que sean emociones bien diferenciadas. Por ejemplo, Bisquerra distingue entre alegría y felicidad; pensamos que conseguir distinguir entre emociones tan similares mediante un formalismo computacional es excesivamente ambicioso, al menos en el estadio actual del proyecto. Así, a los alumnos se les anima a que expresen emociones relativas al miedo, la alegría y la tristeza.

La emoción ira, según bastantes autores, junto al miedo, alegría y tristeza forman las llamadas emociones primarias, las cuales están biológica y genéticamente predeterminadas (son innatas), a diferencia de las emociones secundarias (amor, sorpresa, aversión, vergüenza) que son el producto de aprendizajes, es decir, de combinaciones y modulaciones de las emociones básicas que han sido experimentadas y aprendidas por la persona.

En este estudio inicial se ha dejado la ira fuera para



Fig. 1. Imagen utilizada para recoger textos relativos a la emoción alegría.



Fig. 2. Imagen utilizada para recoger textos relativos a la emoción miedo.



Fig. 3. Imagen utilizada para recoger textos relativos a la emoción tristeza.

abordar el problema a tratar con una mayor sencillez. En futuros trabajos de ampliación del corpus, se prevé incluirla, así como alguna otra emoción si se considere oportuno.

## V. ACTIVIDADES PARA GENERAR EL CORPUS

A continuación se presentan las actividades, adaptadas de [7] (Renom i Plana, 2008), utilizadas para generar los textos que forman el corpus.

### A. Actividad 1: ¿Qué se siente?

- Objetivos:
  - Expresar emociones a través de una imagen dada. (Ver fotografías utilizadas: Figura 1, 2 y 3).
  - Comunicar y describir verbalmente las propias emociones.
  - Diferenciar emociones en situaciones similares.
- Procedimiento: Se le muestran al alumno una serie de fotografías y se le pide que anote la primera emoción que haya sentido y que desarrollen lo que piensan.
- Ejemplo de fotografías:
  - Alegría: Una madre jugando con dos niños y un perro, personas riéndose...
  - Miedo: Personas con la boca abierta y caras de terror.
  - Tristeza: Una persona llorando con ropa sucia y rota.

Variante: Se pone a la vista del usuario bastantes fotografías y se le pide que seleccione la que le causa alguna emoción. A continuación tendrá que explicar por qué la ha seleccionado y qué emoción ha provocado en él.

### B. Actividad 2: Las emociones de la semana/mes

- Objetivos:
  - Recordar situaciones vividas.
  - Descubrir y comparar emociones propias.
- Procedimiento: Se le pide al alumno que recuerde situaciones vividas durante la última semana/mes. A continuación se le pide que escriba por separado aquellas situaciones que tengan que ver con alegría, miedo y tristeza.

## VI. CREACIÓN DEL CORPUS

Cada alumno tuvo que realizar las dos actividades explicadas en el apartado anterior. En el caso de las aulas de España se supervisó personalmente la realización de las actividades para que generaran el mayor número de texto posible. En los países Latinoamericanos se delegó esta responsabilidad en los tutores, cuya implicación fue menor al

ser un apartado voluntario dentro de la Red Internacional e-Culturas y por tanto, de ahí el menor número de textos producidos por estos países.

De las actividades para recoger texto, la Actividad 1, en la que se le pedía al alumno que anotase la emoción que había sentido y lo que había pensado al ver una fotografía, no resultó de utilidad, pues el alumno se limitaba a describir lo que veía en la fotografía sin llegar a expresar ninguna emoción en el texto. Por ejemplo, en la figura 2 se limitaban a comentar que les parecía graciosa la situación, en vez de intentar ponerse en el pellejo de los personajes para describir que estaban asustados, o en la figura 3 que comentaban simplemente “esta triste”, sin llegar a generar más texto sobre el tema...

La Actividad 2, en la que se les pedía que contaran situaciones vividas de cada emoción, sí funcionó perfectamente para recoger texto con contenido emocional. A cada alumno se le pidió que contará 2 situaciones vividas por emoción, de esta forma se llegaron a recoger 1160 textos de alumnos de los 5 países expresando las 3 emociones seleccionadas, tal como muestra la Tabla III. Aunque se ha insistido en que escribieran textos largos, debido a la edad de los alumnos ha sido bastante difícil esta cuestión, obteniéndose una media de 27 palabras por texto, como se puede ver en la Tabla IV. Cabe destacar que para la emoción positiva (alegría) la media de palabras por texto aumenta, así como el número de textos escritos por los niños. Al ser estas actividades emocionales voluntarias, dentro de la Red Internacional e-Culturas, no todos los participantes las realizaron. En la Tabla V se detalla el número de niños por país que las llevaron a cabo.

#### A. Ejemplos de textos recogidos

Los textos que se muestran a continuación han sido escritos por los niños tal y como se muestran, incluyendo las faltas de ortografía.

- Alegría: “El día mas feliz de mi vida fue cuando fui a isla mágica me lo pase muy bien el día mas divertido de mi vida fue ese. Me monte en casi todo y me lo pase muy bien. Ese mismo día fui también a ver un partido de fútbol que también me lo pase muy bien mi equipo empató y empezó perdiendo pero pudieron empatar”.
- Miedo: “Un día estaba en mi casa yo sola porque mis padres se habían ido de viaje y me puse en mi ordenador y sentí un ruido que me asuste y yo me pregunto¿que pasara? y era mi hermanillo que no se había ido y estaba allí en mi casa jugando”.
- Tristeza: “El año pasado estaba yo estudiando para un exámen muy importante, todos los días estudiando un ratito, pero el día del exámen llego, como ya he dicho el exámen era importante; era importante porque de ese exámen dependía si tenia que venir el año siguiente a recuperarlo. Cuando la maestra me dio el

TABLA III  
TEXTOS RECOGIDOS POR EMOCIÓN Y PAÍS

País	Alegría	Tristeza	Miedo	Total
Argentina	66	52	45	163
Brasil	44	41	41	126
Chile	26	22	22	70
España	275	253	239	767
Paraguay	20	8	6	34
<b>Total</b>	<b>431</b>	<b>376</b>	<b>353</b>	<b>1160</b>

TABLA IV  
PALABRAS POR TEXTO

	Alegría	Tristeza	Miedo	Total
<b>Número palabras</b>	14113	9141	8068	<b>31322</b>
<b>Número textos</b>	431	376	353	1160
<b>Media de palabras por texto</b>	32,74	24,31	22,85	27

TABLA V  
NIÑOS PARTICIPANTES POR PAÍS

País	Niños participantes
Argentina	33
Brasil	22
Chile	13
España	139
Paraguay	10
<b>Total niños</b>	<b>217</b>

exámen me puse tan nervioso que la mayoría de las cosas se me olvidaron pero al final cateé el exámen y tuve que recuperarlo”.

#### B. Limpieza de textos

Algunos de los textos generados están en portugués, debido a la participación en la Red e-Culturas de alumnos brasileños. Al ser un porcentaje pequeño la opción tomada fue eliminarlos directamente, para quedarnos así con los textos en español únicamente.

Otro tipo de textos que se eliminaron fueron los generados por aquellos alumnos que no se tomaban en serio la actividad y rellenaban los campos con información que no tenía nada que ver con lo que se le pedía o simplemente se limitaban a escribir caracteres para rellenar texto (“sdfsdfsdfsdfsdf...”).

#### C. Corrector de faltas de ortografía

Una vez disponibles los textos relevantes, se procedió a su corrección ortográfica mediante un sencillo sistema que se explicará más adelante. Lo primero, fue hacer un estudio sobre gran parte del corpus de las faltas de ortografía más típicas que comenten los niños de estas edades al escribir. Algunos errores encontrados:

- Cambiar b por v y viceversa.

TABLA VI  
ALGUNAS PALABRAS SEMILLA

Emoción	Palabras Semilla
Alegría	Ganar, contento, regalar, reír, aprobar, familia
Tristeza	Llorar, pena, lástima, perder, suspender, morir, romper, enfadar
Miedo	Susto, secuestrar, terror, pegar, pesadilla, asustar, inquietud, morir, horror, atacar

- Cambiar c por s y viceversa.
- Cambiar z por s y viceversa.
- Cambiar s por x.
- Cambiar n por m.
- Cambiar i por y.
- Cambiar r por l.
- Omitir la letra h.
- Omitir signos de acentuación.

Para su corrección se ha utilizado un algoritmo que busca en un listado de palabras en español la palabra a corregir aplicándole recursivamente las reglas anteriores mientras no se encuentre la palabra correcta. Para esto se ha utilizado un modelo de RI (Recuperación de Información) clásico mediante el software Lemur, en el que cada palabra del diccionario LaTeX para el español utilizado [8], mediante un etiquetado especial, ha sido indexada como un documento. La forma de proceder era, dada la palabra escrita por el alumno, se buscaba el documento (palabra) en el cual estaban sus letras (equivalentes a las palabras en un sistema clásico de RI).

## VII. EXPERIMENTOS

Se han realizado numerosos experimentos teniendo en cuenta diversas características como la ocurrencia y frecuencia de palabras, n-gramas, etiquetas POS (Part Of Speech) y lemas; el tamaño en Bytes y número de palabras del texto completo y la media por frase; ocurrencia y frecuencia de trigger words, palabras enfatizadas, símbolos especiales... En este trabajo las trigger words están formadas por conjuntos de unas 10-15 palabras semilla que se han seleccionado manualmente a partir del corpus de entrenamiento (véase Tabla VI). Algunos de los experimentos en los que se han obtenido mejores resultados son los siguientes:

- a. SEM: Ocurrencia de las palabras semilla.
- b. DIC: Ocurrencia de las palabras de diccionario.
- c. NGR: Ocurrencia de n-gramas de más de 4 letras.
- d. DIC+POS: Ocurrencia de las palabras de diccionario y etiquetas POS.
- e. NGR+POS: Ocurrencia de n-gramas de más de 4 letras y etiquetas POS.
- f. FDIC+POS: Frecuencia de las palabras de diccionario y etiquetas POS.

TABLA VII  
RESULTADOS DE LA CLASIFICACIÓN DE EMOCIONES

Exp.	Emoción	Accuracy	Precisión/ cobertura
SEM	Alegría	0,796	0,681/0,833
	Miedo	0,918	1,000/0,692
	Tristeza	0,755	0,687/0,611
DIC	Alegría	0,829	0,844/0,675
	Miedo	0,914	0,893/0,807
	Tristeza	0,762	0,765/0,382
NGR	Alegría	0,800	0,771/0,675
	Miedo	0,905	0,862/0,807
	Tristeza	0,800	0,809/0,500
DIC+	Alegría	0,839	0,849/0,700
	Miedo	0,905	0,862/0,807
POS	Tristeza	0,771	0,778/0,412
	Alegría	0,819	0,800/0,700
NGR+	Miedo	0,895	0,857/0,774
	Tristeza	0,781	0,789/0,441
FDIC	Alegría	0,815	0,831/0,592
	Miedo	0,835	0,968/0,539
POS	Tristeza	0,792	0,812/0,415
	Alegría	0,770	0,803/0,523
FSEM+	Miedo	0,727	0,976/0,276
	Tristeza	0,814	0,728/0,357

- g. FSEM+POS: Frecuencia de las palabras semilla y etiquetas POS.

Finalmente, se ha procedido a entrenar un algoritmo de aprendizaje supervisado con el corpus de entrenamiento, evaluando la bondad del modelo sobre el corpus de evaluación para cada uno de los experimentos anteriores. El algoritmo de aprendizaje automático utilizado en esta experimentación es SVM, [9] (Cristianini & Shawe-Taylor, 2000).

## VIII. ANÁLISIS DE LOS RESULTADOS

Los resultados obtenidos se describen en la Tabla VII. A pesar del sencillo modelo de categorización implementado los resultados son esperanzadores, si bien hay diferencias significativas entre cada emoción dependiendo del experimento realizado, por ejemplo, tanto en NGR y NGR+POS mejoran los resultados obtenidos para tristeza y empeoran los de miedo y alegría si los comparamos con sus experimentos predecesores DIC y DIC+POS respectivamente.

Si en vez de utilizar simplemente la ocurrencia de las características, como se hace en DIC+POS, se tiene en cuenta la frecuencia de las mismas, tal es el caso de FDIC+POS, se vuelve a obtener resultados mejores para tristeza, empeorándose ligeramente los de alegría y miedo.

Utilizando solamente la ocurrencia de las palabras semilla, SEM, en general se consiguen mejores resultados que midiendo la frecuencia de las mismas, FSEM+POS, por tanto en las siguientes líneas y debido a que es el caso con más margen de mejora, se tratará de forma más exhaustiva este tema haciendo un estudio de los errores cometidos.



Como es usual en las tareas de categorización, una precisión elevada se suele conseguir penalizando la cobertura, tal es el caso del miedo en SEM con una precisión de 1, pero a costa de una discreta cobertura que no alcanza el 70% de los textos marcados con tal emoción. Sin duda, este resultado nos indica que el conjunto de palabras semilla seleccionado para esa emoción es preciso, pero incompleto. En el otro extremo se encuentra la alegría, que obtiene una cobertura que supera el 80%, pero con pobre resultado en cuanto a la precisión: un 0,68. Esto es, posiblemente superar esa cobertura a partir de las sencillas características léxicas y morfosintácticas utilizadas sea muy difícil, si no imposible, pero sí pensamos que es posible reducir la lista de palabras semilla sin que la cobertura se resienta. Finalmente, la tristeza se muestra como la emoción más difícil, con un resultado discreto tanto en precisión como en cobertura.

En la Tabla VIII y Tabla IX se muestran algunos errores cometidos por el categorizador en lo relativo a la categoría de tristeza, debido a que aparecen palabras semilla en una categoría distinta a la que se le presupone. Tal es el caso de los textos 1 y 3 de la Tabla VIII, ambos errores inducidos por la palabra ganar que se corresponde con la categoría alegría. Similar situación, pero a la inversa se da en los textos 1 y 2 de la Tabla IX y el verbo perder. Al margen que estos casos concretos indican que ganar/perder requiere un tratamiento particular, parece claro que para corregir estos casos se requiere un análisis sintáctico más profundo. Otros motivos de error son escritos con un vocabulario muy aséptico, aunque lo que allí se describa tenga una fuerte carga emocional, tal es el caso del texto 2 de la Tabla VIII.

Un caso especialmente interesante es el último texto de la Tabla IX: el categorizador marca con la categoría miedo un texto que el niño escribió como respuesta a una actividad pensada para que éste manifieste tristeza. Sin embargo, el texto claramente relata una situación en la que el niño pasó miedo, pero la recuerda con tristeza, no con miedo. Esto nos indica, por una parte, que los niños a estas edades pueden confundir ambos sentimientos y que puede darse solapamiento. En consecuencia es necesario etiquetar manualmente cada texto y contemplar la posibilidad de que haya textos pertenecientes a más de una categoría.

## IX. CONCLUSIONES Y TRABAJO FUTURO

Se ha presentado aquí la creación de corpus para el idioma español a partir de textos escritos por niños de 10 y 11 años relativos a experiencias en las que hayan sentido miedo, tristeza o alegría. Hemos descrito la metodología seguida para la creación del corpus. Para validar el corpus resultante se ha procedido a realizar una experimentación preliminar, orientada a la categorización automática de los sentimientos. Los resultados que hemos obtenido son en buena medida consistentes con el corpus tal cual está etiquetado manualmente, alcanzando una precisión media de 82.7% para una cobertura del 60%. El análisis de errores realizado por el categorizador desarrollado muestra que el enfoque para crear el corpus es el adecuado: en buena medida los

TABLA VIII  
EJEMPLOS DE FALSOS NEGATIVOS DE LA CATEGORÍA TRISTEZA. LOS TEXTOS SE PRESENTAN TAL CUAL FUERON ESCRITOS POR LOS NIÑOS

### Falsos negativos de la emoción tristeza

Que el primer año de moto cros quede segundo por un puntito y no pude **ganar el campeonato**.

El año pasado estábamos toda mi familia en el campo de mi tío en una matanza. Estaba toda mi familia y cuando llegamos fuimos a matar lo cerdos. Los matamos y cuando terminamos dijo mi primo yo voy a echarle de comer a los caballos y mi padre y yo nos fuimos con él porque nos gustan mucho. Estábamos llegando a la cuadra cuando dice mi primo qué raro que hoy no haya relinchao el caballo porque siempre que vengo por aquí le silvo y él relincha. Cuando llegamos a la cuadra no lo veíamos; abrimos la puerta y nos lo encontramos en el suelo tirado y estaba muy hinchado. Mi padre lo primero que hizo fue coger un chaleco mojarlo y ponérselo en la cabeza y parecía que había reaccionado pero llamamos al veterinario para que lo viera. Llegó en un momento le dijimos lo que había pasado y le puso un pinchazo para dormirlo; lo estuvo observando y le había dado un cólico y si hubiéramos llegado mas tarde hubiera muerto. Le pusimos oxígeno y suero y el caballo se levantó. Al día siguiente el caballo ya estaba mucho mejor y **ahora mi primo se monta en él**.

Ache 1 mes icimos un campeonato en educacion fisica, nos decepcionamos por que perdimos. Estabamos jugando el A contra el B nosotros que somos el b, perdimos gano el a porque nosotros estavamos muy cansados pero esque si no, ese día ganamos el partido encima con la calor. Que acia que no iba a perder y quedars e agotado.

TABLA IX  
EJEMPLOS DE FALSOS POSITIVOS. ENTRE PARÉNTESIS LA CATEGORÍA ASIGNADA POR EL CATEGORIZADOR. LOS TEXTOS SE PRESENTAN TAL CUAL FUERON ESCRITOS POR LOS NIÑOS

### Falsos positivos

(**alegría**) El día mas feliz de mi vida fue cuando fui a isla mágica me lo pase muy bien el día mas divertido de mi vida fue ese. Me monte en casi todo y me lo pase muy bien. Ese mismo día fui también a ver un partido de fútbol que también me lo pase muy bien mi equipo empate y empezó perdiendo pero pudieron empatar  
\*\*\*

(**alegría**) Mi primo un día fue al Rocío con migo y cuando estábamos allí nos encontramos una cartera que tenia 500\$, tarjetas de créditos y una foto de su familia y lo entregamos a objetos perdidos pero los 500\$ los repartimos entre los dos y toco a cada uno 250\$. **Que suerte encontrar tanto dinero.** (**alegría**)

(**miedo**) pues mira a mí me da tristeza cuando la gente abanandonan a los animales , cuando alguien se muere o esta malita , cuando perdemos un partido de futbol , cuando me castigan en mi clase , cuando los reyes magos no me regalan na, cuando las niñas me ignoran.

(**miedo**) me perdi en el supermercado y me largue a llorar pero del miedo cuando tenia 4 años

errores tipo cometidos requieren ampliar el conjunto de palabras semillas o añadir algunas características basadas en reglas morfo-sintácticas. Sin embargo, algunos errores como la distinción entre miedo y tristeza son más sutiles, y posiblemente requiera de técnicas más afinadas, como el tratamiento individualizado para algunas palabras, tal es el caso de los verbos ganar/perder.

En lo relativo a la creación del corpus, éste aún tiene un tamaño reducido, pero esperamos que a lo largo del curso escolar su tamaño se incremente muy significativamente, no sólo porque los niños sigan haciendo uso de las actividades ya diseñadas, sino también por la inclusión de nuevas

TABLA X  
PREVISIÓN DE PARTICIPACIÓN EN LA EXPERIENCIA E-CULTURAS PARA ESTE  
AÑO ESCOLAR Y SU INCIDENCIA EN LA AMPLIACIÓN DEL CORPUS

País	Niños participantes	Documentos por emoción			Total
		Alegría	Tristeza	Miedo	
Argentina	19	38	38	38	114
Brasil	14	28	28	28	84
Chile	69	138	138	138	414
España	125	250	250	250	750
Paraguay	128	256	256	256	768
Total	355	710	710	710	2130

actividades y la ampliación a un mayor número de alumnos con acceso a la plataforma de e-Culturas. Con esto, la estimación para este año se describe en la Tabla X, donde 355 nuevos alumnos generarán unos 2130 documentos con una media de unas 30 palabras por texto.

### BIBLIOGRAFÍA

- [1] B. Pang y L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, 2 (1-2), 1-135, 2008.
- [2] R. Bisquerra, "Educación Emocional y Bienestar," Barcelona: Praxis, 2002.
- [3] A. Pantoja, "Internet y la construcción de una ciudadanía intercultural. Balance de una experiencia," I Jornadas Internacionales y VI Jornadas sobre Diagnóstico y Orientación El carácter universal de la educación intercultural, pp. 72-128, 2008.
- [4] M. Alcaide, E. Blanco, A. Pantoja y A. Jiménez, "Capacitación de maestros en valores interculturales a través de la Red Internacional e-Culturas," INECE'08 Madrid, 2008.
- [5] e-Culturas. Recuperado el 15 de Enero de 2010 de <http://www.e-culturas.org>
- [6] D. Goleman, "Inteligencia emocional: a teoría revolucionaria que define que é ser inteligente," (M. Santarríta, Trad), Rio de Janeiro: Objetiva, 2001.
- [7] Renom i A. Plana, "Educación Emocional: programa para educación primaria (6-12 años)," Wolters Kluwer España, 2008.
- [8] Diccionario LaTeX para el español. Recuperado el 15 de Enero de 2010 de <http://copa.uniandes.edu.co/software/latex/manual.html>
- [9] N. Cristianini y J. Shawe-Taylor, "An Introduction to Support Vector Machines and other kernel-based learning methods," Cambridge University Press, 2000.
- [10] T. O'Reilly, "What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software", 2005.
- [11] M. Quarterly, "How companies are benefiting from Web 2.0. techreport", 2009.
- [12] A. Esuli & F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource". Proceedings of LREC 2006 - 5th Conference on Language., 2006.
- [13] Y. Seki, D. K. Evans, L.-W. Ku, H.-H. Chen, N. Kando & C.-Y. Lin, "Overview of Opinion Analysis Pilot Task at NTCIR-6", (págs. 265-278), 2007.
- [14] I. Ounis, C. MacDonald, M. de Rijke, G. Mishne & I. Soboroff, "Overview of the TREC 2006 Blog Track. Special Publication 500-272". National Institute of Standards and Technology (NIST), 2006.
- [15] I. Ounis, C. Macdonald & I. Soboroff, "On the TREC Blog Track", 2008.
- [16] A. Passant & P. Laublet, "Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data", 2008.



**Eladio Blanco López**, Ingeniero en Informática por la Universidad de Jaén, se encuentra acualmente realizando el Doctorado en Informática en la misma universidad. Miembro del grupo de investigación IDEO desde 2008. Su área de investigación es el Procesamiento del

Lenguaje Natural, Análisis de Sentimientos y Minería de Opinión.



**Fernando Martínez-Santiago** is full professor of Databases at the Technical School. He is a member of the SINAI group in the Department of Software and Computing Systems at the University of Jaén. He received his Master's degree in Computer Science at the University of Granada, Spain and his PhD in Computer Science at the UNED, Spain. His research interests are cross-lingual information retrieval,

semantic analysis, specifically textual entailment and subjective language analysis, and its application to other NLP areas, such as Question Answering and Sentiment Analysis. He is the author of more than 20 research papers for international journals and conferences. He has been a reviewer and a member of many program committees of international conferences and has been involved in different national and international research projects.



**Antonio Pantoja Vallejo**, Profesor Titular de Universidad de Métodos de Investigación y Diagnóstico en Educación - Universidad de Jaén - España. Facultad de Humanidades y CC. de la Educación - Departamento de Pedagogía. Doctor en Filosofía y Ciencias de la Educación por la UNED con premio extraordinario de doctorado. Profesor de la Universidad de Jaén desde 1998, titular del área de Métodos de Investigación y Diagnóstico en Educación desde 2002. Coordinador

de varios programas de doctorado nacionales como internacionales, que en la actualidad están en fase de tesis doctorales. Ha dirigido numerosas investigaciones que han dado lugar a la obtención del Diploma de Estudios Avanzados (DEA), así como diversas tesis doctorales vinculadas con mis líneas de investigación. Es el responsable del Grupo de Investigación IDEO (HUM 660) de la Junta de Andalucía (<http://www.grupoideo.net>) y coordinador de la "Red internacional e-Culturas" (<http://www.e-culturas.org>). Director académico del Máster en Educación que organizan conjuntamente la Universidad de Jaén y FUNIBER. Ha participado como ponente y conferenciante en congresos, jornadas y seminarios nacionales e internacionales. Posee dos tramos de investigación del CNEAI. Autor de diversos libros, capítulos de libros y más de una treintena de artículos en revistas científicas. Ha dirigido una docena de tesis doctorales.



**L. Alfonso Ureña-López** is Assistant Professor in the Department of Computer Science at Jaén University (Spain). He received M.S. degree in Computer Science from the University of Granada, and Ph. D. in Computer Science from Software Engineering Department of Granada. His Ph. D. Thesis was the winner of the 2001 Awards of the Spanish Society for Natural Language Processing. His current main research interest is Intelligent Search, including

Search Engines, Search Assistants, Natural Language Processing Tools for Text Mining and Retrieval and Human Computer Interaction. Relevant topics include Multilingual-Multimodal Information Access, Information Synthesis and Summarization, semantic networks (wordnets, lexical acquisition, web as corpus, word sense disambiguation), web search engines (clustering and visualization of search results, portal search engines).

Dr. Ureña was Director of the Computer Science Department at University of Jaén (1997-2004). Currently he is Deputy Director of Polytechnic School at Jaén. He is Director and founder of Research Group of Intelligent Systems of Information Access. Also, he is President of SEPLN from 2007 and editor of *Procesamiento de Lenguaje Natural Journal*. He has given several scientific conferences, both in Spain and abroad. He has directed several research projects and several contracts with companies in technology transfer. He has directed several Ph. D. in Computer Science and research works. He has participated in numerous courts dissertations at universities in Spain and abroad. Also, He has directed and participated in numerous courses at several universities. He is author or co-author of more than 100 scientific publications. He is technical reviewer in several journals and in the Program Committee of some major conferences.