

Desarrollo de una herramienta informática de análisis y clasificación de textos auxiliar de investigaciones sobre procesos de alfabetización

Facultad Regional Rosario - UTN

Facultad Regional Santa Fe - UTN

Zeballos 1341- Rosario

kantor@frr.utn.edu.ar

Resumen

El proyecto en su plan original se proponía desarrollar una herramienta informática auxiliar de actividades de investigación vinculadas con los procesos de enseñanza-aprendizaje de lectoescritura. La propuesta se encuadraba, en el marco del llamado "desarrollo experimental" consistiendo en la realización de trabajos sistemáticos para la producción de un nuevo producto basado en conocimientos existentes y disponibles. Elemento fundamental en la concepción del proyecto es la permanente interacción con los potenciales usuarios: grupos de investigación dedicados a la lectoescritura, dado que el objetivo principal es que las realizaciones sean útiles a la comunidad. Este proceso de realimentación positiva, hizo que los objetivos iniciales se fueran modificando a la luz de las observaciones y solicitudes.

Así, actualmente se están desarrollando dos módulos, uno capaz de analizar un texto simple (que debe ser incorporado con un mínimo de preprocesamiento por parte del investigador) produciendo un texto asociado donde cada unidad de análisis va acompañada de renglones informativos y otro módulo que permitirá estudios estadísticos vinculados con los temas a investigar determinados por el operador.

Se contará con una amigable interfase y será desarrollado para diversos sistemas operativos, con una amplia gama de procesadores de texto.

Se realiza bajo los criterios del software libre, tanto en las herramientas a utilizar, como a la disponibilidad pública de los resultados.

Se propone experimentar con una modalidad de producción, mediante una página web actualizada según avances, permitiendo aportes externos, tanto en la construcción del código como en la definición de las especificaciones de los productos.

Palabras clave: alfabetización, lecto-escritura, , software libre.

Contexto

El proyecto está radicado en la Facultad Regional Rosario de la UTN y se realiza conjuntamente con docentes de la Facultad Regional Santa Fe.

Está acreditado en la UTN, con vencimiento el 31/12/2011.

En este momento está en trámite la formalización de la relación con los investigadores del proyecto "Los niños sordos y oyentes como editores: los procesos de revisión y comprensión lectora" dirigido por la Dra. Mónica Baez, de la Universidad Nacional de San Miguel (FGR-sede Rosario), por medio de un convenio.

Introducción:

El principal objetivo es la construcción de herramientas que sean de uso efectivo por los investigadores en sus actividades concretas, con prestaciones y calidad similar al de los productos del mercado.

Se estima que, como efecto paralelo, la experiencia sobre la aplicación del modo de desarrollo propuesto (funcionando en modo interactivo con la comunidad potencial de usuarios y abiertos a la colaboración técnica de participantes externos) producirá resultados interesantes desde el punto de vista metodológico.

Se intenta efectivizar la construcción de estos productos bajo la modalidad del llamado software libre.

Es entonces que la propuesta se desarrolla en código abierto, sin reservar más derecho que el de que no se pueda, de ninguna manera, establecer limitaciones a su disposición pública.

Se ofrecen públicamente los progresos del proyecto, poniendo a disposición de los posibles usuarios del producto cada uno de los logros conseguidos, junto con el código abierto de los programas para quienes les

interese conocer y eventualmente participar en el aspecto de la elaboración técnica subyacente.

De esta manera se brinda la posibilidad, a quienes quieran aportar ideas, de incorporar elementos desde el punto de vista técnico de los productos, como en cuanto a sugerir mejoras o agregados a las prestaciones, manteniendo una conducción centralizada de versiones y documentación.

La propuesta se basa en la interacción de dinámica entre nuestro grupo de trabajo, que toma a su cargo la construcción de las herramientas, y un grupo formado por potenciales usuarios que denominamos Comité de Usuarios (C.U.).

La principal función del C.U. es probar las versiones del proyecto, justamente, como usuarios, proponiendo mejoras. Otra misión del mencionado comité es la de considerar las propuestas y sugerencias de quienes, no integrando el C.U., utilicen versiones de prueba del producto, y a partir de su empleo propongan modificaciones. Es el C.U. quien juzga esas propuestas tanto sobre la incorporación de mejoras y nuevas prestaciones, como a los aspectos de tipo ergonómico como colores, diseño, usabilidad, etc.. A partir de la publicación de los programas fuente se reciben aportes de desarrolladores externos. En ese sentido nuestro grupo se reserva la responsabilidad del análisis y consideración de las propuestas manteniendo el control de versiones.

De cada módulo en desarrollo se ofrecerán versiones para las plataformas Linux y Windows, con idénticas prestaciones y características funcionales, por lo que podrán ser usadas prácticamente en cualquier equipo.

El primer módulo en desarrollo es el analizador de textos.

Este módulo toma un texto (que puede ser escrito con diversos procesadores de texto libres o propietarios) con un mínimo de preprocesamiento (sólo se deberán indicar cuales son para el investigador las unidades de análisis). Esto facilitará el trabajo de transcripción.

Eventualmente se utilizan como fuente escritos realizados por los niños directamente en la computadora, con la única exigencia de delimitar las unidades conceptuales.

En simultáneo con el desarrollo del analizador de textos, se construyen

herramientas orientadas al procesamiento estadísticos de los resultados.

Se ha seleccionado como lenguaje programación a utilizar el denominado Python y se ha elegido MySQL para la base de datos,

Resultado Obtenidos/Esperados

No hemos podido encontrar en la información relevada en nuestra actividad preliminar trabajos actuales o recientes con similares orientaciones a los de nuestro proyecto.

Los analizadores de texto relevados sólo cumplen parcialmente con las algunas de las funcionalidades que ofrecerá nuestro proyecto y, muy importante, en todos los caso se tratan de productos de distribución no-libre

Como principal antecedente podemos mencionar lo realizado por la Prof. Isabel García Hidalgo de la Universidad de México alrededor de 1994, que fue utilizado durante muchos años. Sin embargo, con el paso del tiempo, el notable avance de la tecnología informática hizo que esta creación denominada TEXTUS fuera, poco a poco, quedando obsoleta ya que funcionaba en ambiente DOS y se debía para su uso utilizar complejos mensajes de texto indicando comandos y parámetros. El producto exigía además muy estrictas especificaciones respecto al preprocesamiento de los textos a analizar y para las actividades de análisis estadístico solo trabajaba sobre una ya hace tiempo desactualizada base de datos.

Las herramientas en desarrollo colaboraran en la delimitación, análisis e interpretación de datos según intereses específicos inherentes a la investigación de índole interdisciplinar e intradisciplinar, en y desde campos disciplinares fundados en la interpretación de textos escritos y procesos de textualización escrita.

En este sentido, entendemos que se trata de una contribución importante para la indagación psicolingüística focalizada en la comprensión de los procesos de alfabetización.

Contar con herramientas informáticas de este tenor, flexibles, versátiles para en el ámbito académico, educativo y clínico, además de promover el diálogo multidisciplinar contribuirán también a la revisión y problematización de parámetros teóricos vinculados tanto a las prácticas

propias de esos ámbitos como los relativos a la determinación de “normalidad “ y “patología” en niños, jóvenes o adultos en proceso de alfabetización.

Formación de Recursos Humanos

Las actividades propuestas contribuirán directamente a la formación de los integrantes del proyecto, así como de quienes participen en el mismo dada su modalidad de construcción abierta.

En particular se debe remarcar el impacto en la formación de los integrantes de las cátedras de Algoritmos y Estructuras de Datos que participan en el proyecto.

Adicionalmente se espera que la adopción de las herramientas auxilien positivamente las actividades de investigación de los procesos de lectoescritura y, por carácter transitivo, constituya una efectiva contribución a la formación de quienes son el objetivo final de estos trabajos : las personas en proceso de alfabetización.

REFERENCIA:

Proyecto de Investigación: “Los niños sordos y oyentes como editores: los procesos de revisión y comprensión lectora”

Unidad académica: Escuela de Humanidades

Unidad de investigación: Licenciatura en Psicopedagogía

UNSAM/FGR-sede Rosario

BIBLIOGRAFIA

“Caperucita Roja aprende a escribir”
Estudios comparativos en tres lenguas.

Ferreiro, E., Pontecorvo, C., Ribeiro Moreira, N. García Hidalgo, I.,
Barcelona, Gedisa, 1996, 269 pp

"The TEXTUS system", Theories, models and methodology: Current trends in research on writing.

García Hidalgo, María Isabel

Eds. Rijlaasdam, Huub van den Bergh y Michel Couzijn. Amsterdam University Press, The Hague, 1995.

Tesis: “Recursos Gráficos y Discursivos Empleados Por Niños En La Edición De Tres Textos Específicos” Magister Mónica O. Baez Flamini – Directora: Dra- Emilia Ferreiro Schiavi(2009)

*Python Reference Manual “ Guido Van Rossum

*Core Python Programming - Wesley J. Chun -Publisher: Prentice Hall PTR

First Edition December 14, 2000 - ISBN: 0-13-026036-3

Analizador Sintáctico: python rdparser.py

www.plonehispano.com/

plone.org/countries/conosur

.