

El lenguaje natural en plataformas de educación a distancia.

Carina FRACCHIA **Sandra ROGER**
cfracchi@uncoma.edu.ar sroger@uncoma.edu.ar

Departamento de Ciencias de la Computación
Universidad Nacional del Comahue
Buenos Aires 1400-(8300) Neuquén, Argentina.
FAX: (54)(0299)4490300 int. 435

Palabras Claves:

Informática Educativa, Educación a Distancia, Resolución de Problemas, Lenguaje Natural, HCI.

Resumen

Siguiendo nuestra línea de investigación y plataforma propuesta para el dictado de la materia Resolución de Problemas y Algoritmos (RPA) a distancia, hemos apuntando, dentro del ámbito HCI (Interacción Hombre-Maquina), al desarrollo de un entorno inteligente, el cual hace énfasis en el uso del procesamiento del lenguaje natural. De esta manera, se tenderá a que la interacción del alumno-profesor-sistema se desarrolle en forma natural, logrando cumplir con los principios de una buena interfaz-usuaria(amigable, simple, fácil de usar, etc.).

1 Introducción

La educación a distancia requiere de paradigmas distintos a los de la educación tradicional. Actualmente, cuenta con una plétora de desafíos en los cuales los investigadores han enfocado sus trabajos.

Profundizando el diseño de la interfase-usuario para lograr una buena comunicación con el usuario (profesor-alumno), hemos centrado nuestros esfuerzos en la priorización del estilo de interacción lenguaje natural, para responder a las siguiente clase de preguntas: ¿Cómo el lenguaje natural puede ser insertado en las diferentes actividades de una curricula? ¿Para qué y que clase de beneficios obtenemos al introducir el lenguaje natural en la educación a distancia? entre otras.

La interfaz en lenguaje natural es quizás, el sueño y el ideal de los usuarios inexpertos en diferentes ámbitos, inclusive fuera de la educación, ya que permite la interacción con la computadora o algún otro medio electrónico haciendo uso un lenguaje común o “natural”. No se requieren habilidades especiales del usuario cuando se interacciona utilizando lenguaje natural. De esta manera podemos tratar con diferentes niveles de alumnos, según la clasificación dada en [FR02b].

Las interfaces en lenguaje natural están en sus albores, y las aplicaciones particulares existentes, donde otro tipo de interfaz sería poco conveniente, son intentos limitados. Los problemas de implantación y la demanda extraordinaria en recursos de cálculo han limitado el desarrollo de las interfaces en lenguaje natural. Sin embargo, numerosos programadores e investigadores están trabajando de manera activa en el desarrollo de interfaces en lenguaje natural, de tal forma que es un área que definitivamente continuará creciendo y por ello requerirá de un estrecho seguimiento.

Para el presente trabajo se realizaron investigaciones sobre recursos y algoritmos básicos de tecnologías que hacen uso del lenguaje y que permitan ser utilizadas para mejorar el acceso a información textual multilingüe. Esto se debe a que el material didáctico puede presentarse, en nuestro caso, tanto en español como en inglés.

En el ambiente propuesto en [FR02b] se presentan herramientas de comunicación (correo, chat, foros, etc.) y se desarrollan herramientas soporte de contenido y de metodología de trabajo (tutoriales, generador de llaves, auto-evaluador, etc.) y herramientas para el seguimiento del aprendizaje.

Se pretende desarrollar un sistema inteligente que brinde:

- Soporte de comunicación y colaboración donde, tanto el docente como el alumno, podrán expresar verbalmente instrucciones relativas a la disponibilidad del material didáctico y a otras personas, mandarles mensajes, consultar por la presencia de alguna personas, realización de alguna actividad, etc.
- Acceso a la información. Proporcionada mediante el acceso verbal a fuentes de información dada por la cátedra o la web.
- Un buscador de noticias en Internet capaz de interrelacionar búsquedas en cualquiera de los idiomas trabajados.
- Un soporte capaz que relacione documentos y consultas con independencia del idioma en que fueron formulados originalmente.

Un ejemplo típico en la literatura de procesamiento de lenguaje natural para la ambigüedad, es la siguiente oración

Juan vio un mono con un telescopio.

- Aceptación 1: El mono tenía un telescopio y Jun lo vio.
- Aceptación 2: Haciendo uso de un telescopio, Jun vio a un mono.

Figura 1: Ejemplo de Ambigüedad. Un ejemplo con diferente árboles de derivación

Para ello será necesario conseguir mejoras sustanciales en el área de resolución de la ambigüedad léxica y desarrollar técnicas robustas que permitan, para un documento en cualquiera de los idiomas contemplados, desarrollar los recursos lingüísticos necesarios para abordar estas tareas.

En la sección 2 se fundamenta la selección del estilo de interacción lenguaje natural. En la sección 3 se describe y ejemplifica la implementación del lenguaje natural en las herramientas de soporte de contenido y metodología de trabajo. En la sección 4 se enunciarán las conclusiones y trabajos futuros del presente trabajo.

2 Fundamentos de la utilización del Procesamiento en Lenguaje Natural

El lenguaje es un aspecto fundamental del comportamiento humano y crucial en nuestras vidas. En forma escrita, sirve para registrar y transmitir información de generación a generación. En forma hablada sirve como principal medio de coordinación de nuestro comportamiento.

Los sistemas de lenguaje natural deben usar un conocimiento considerable acerca de la estructura del lenguaje mismo, incluyendo qué es una palabra, como las palabras son combinadas para formar sentencias, que es lo que significan, y como su significado contribuye al de las sentencias. Además de este conocimiento, es necesario el conocimiento del mundo, para poder entablar un diálogo([AHM00]).

Sin embargo, si se utiliza el lenguaje natural para expresarse en el idioma común, las posibilidades de éxito para lograr una comunicación efectiva con la computadora son mínimas. El uso de lenguaje natural presenta una serie de inconvenientes propio del lenguaje natural que se quiera modelar tales como ambigüedad de los sentidos de una frase ([Ped01])(Fig. 1), resoluciones de anáforas y elipses ([Mit01],[PSNn⁺01],[Rog01])(Fig. 2), etc.

Dada la siguiente regla:

Buscar todos los alumnos que realizaron el diagrama 5 con un bucle.

Se observa una serie de acepciones, por ejemplo:

*Ana abre la puerta y la cierra tras de sí.
Ana vive con Jun y cocina para él cada día*

En la primera oración el pronombre *sí* está correferenciando a *Ana*. En el segundo ejemplo *él* lo hace con *Jun*. Es decir que existen dos palabras que referencian a un mismo concepto: un antecedente y una anáfora.

La casa en la playa es muy bonita y Ø fue diseñada por José

A diferencia de la anáfora, la elipse no se coloca ninguna palabra para correferenciar al antecedente.

Figura 2: Ejemplo de Anáfora y Elipse

- Aceptación 1: buscar todos los alumnos que para realizar el diagrama 5 usaron un bucle.
- Aceptación 2: Usando un bucle buscar todos los alumnos que realizaron el diagrama 5.

El procesamiento del lenguaje natural está dividido en los siguientes niveles de conocimientos principales([All95]):

- Fonético y fonológico: está relacionada en cómo las palabras son relacionadas a las sentencias.
- Morfológico: se refiere a cómo las palabras son construidas.
- Sintáctico: se refiere a cómo las palabras son agrupadas para formar sentencias correctas.
- Semántico: en este conocimiento se trata con el significado de las palabras y cómo estos significados son combinados para formar el de la sentencia.
- Análisis pragmático: se refiere a cómo las sentencias son usadas en diferentes situaciones y cómo esto afecta en su interpretación.
- Del discurso: esta relacionado en cómo el significado de una sentencia afecta el de la sentencia siguiente.

Las tres primeras etapas se podrían encuadrar en el reconocimiento de la sentencia que se quiere modelar y las restantes formarían parte de la comprensión de la misma.

3 Herramientas de Soportes de Contenidos y Metodologías de Trabajo

El ambiente propuesto para el dictado de la materia Resolución de Problemas y Algoritmos a distancia ([FR02b]) no se limita a usar la tecnología sólo como medio de comunicación sino que utiliza el potencial de ésta como soporte de construcción común de conocimiento.

El entorno de esta metodología a distancia requiere de cierta flexibilidad en lo que se refiere a tiempo, espacio, ritmo y acceso. El grado de flexibilidad de estos elementos y del conjunto del proceso de formación depende, en gran medida, del tipo de relación entre el docente, los alumnos y los contenidos.

En los ambientes tradicionales el alumno cuenta con una relación social inherente, es por ello que nuestra propuesta tiende a crear un medio por el cual esta característica se preserve. En este sentido es dónde cobra fuerza el lenguaje natural, ya que se flexibiliza la comunicación dejando de ser exclusivamente en forma escrita, permitiendo con ello, que el alumno se sienta más cómodo y reduciendo el tiempo de adaptación.

La interacción en lenguaje natural dentro de este ambiente, es realizada de la siguiente manera: los usuarios (alumno, profesor y sistema) tienen la opción de poder comunicarse, o dialogar, en forma escrita o verbal, si es realizada de esta última forma, se procede mediante expresiones no restringidas en lenguaje natural, es decir, se pretende que la comunicación sea más robusta y que las expresiones no se limiten a un lenguaje de comandos. De esta manera se pretende que:

- Usuario se encuentre dentro de un entorno en el cual se “intuye” el tipo de comunicación que establecerá con el sistema. Es decir, que se el tipo de conversación que el alumno o el docente establezca estará restringido a un dominio específico, en nuestro caso a temas relacionados a la materia y el entorno del sistema.
- El usuario puede establecer múltiples conversaciones con el sistema dentro de un mismo entorno. Estos diálogos siempre se encuentran dentro de algunos de los diálogos que tiene previstos el sistema.
- El sistema cuente con una información de contexto muy rica. La información los diálogos que se están produciendo se almacena para que pueda ser utilizada por los diferentes módulos que componen el sistema, incluido el módulo de lenguaje natural y permitir de esta manera poder manejar cambios de contextos y resoluciones de anáforas, ambigüedades y otros fenómenos lingüísticos asociados al procesamiento del lenguaje natural.

Para el proceso de reconocimiento de voz se pueden desarrollar herramientas para tal tarea o emplear alguna comercial tal como el ViaVoice de IBM (utilizadas en proyectos similares: [AHM00]). En la síntesis de voz se puede también usar la capacidad de algún Text to Speech de Via Voice.

Uno de los principales beneficios del uso del lenguaje natural en plataformas de la educación a distancia es que no se requieren alumnos avanzados o con conocimiento alto en herramientas en computación. Es decir, que los alumnos que se encuentran dentro del nivel bajo (sección 3.1 de [FR02b]) pueden resultar beneficiados en mayor medida con el uso de una interfaz que utilice el lenguaje natural.

Tanto para la tarea de evaluación de conocimientos previos para la materia resolución de problemas y algoritmos como para el seguimiento del alumno, el docente podrá, a través de

consultas simples, seleccionar y recuperar información sobre el material bibliográfico disponible, que considere necesario para enviar o recomendar al alumno.

[FR02a]

Las herramientas propuestas mencionadas en la sección 1 son:

- Entre las herramientas de comunicación
 - **Correo** Tanto los alumnos como los profesores cuentan con una dirección de correo electrónico.
 - **Chat** Se trata de un espacio virtual mediante el cual es posible establecer contacto con distintas personas en tiempo real.
 - **Foros** Espacio para debate e información, ya que pueden participar tanto docente como alumnos.
 - **Calendarios o agendas** Contendrá toda clase de información y avisos sobre la materia y su contenido.
- Entre las herramientas soporte de contenidos y de metodología de trabajo encontraremos:
 - **Tutoriales, Ejercitativos** para los distintos contenidos trabajados en la materia RPA
 - **Generador de Llaves** es una aplicación que se desarrollará en la sección ??.
 - **Auto-Evaluador** el cual sirve para que el alumno vaya autoevaluando sus conocimientos.
 - **Espacio de Debate y Construcción** está pensado para la crítica y participación de los alumnos con respecto a la resolución de ejercicios. La idea básica de este punto es poder brindar al alumno un marco mediante el cual pueda construir sus estructuras cognitivas. Este espacio se presenta como generador de debates, los cuales contribuyen al proceso del desarrollo cognitivo individual.

El potencial de estas herramientas se incrementará si se combinan con el uso del lenguaje natural. Por ejemplo se podrían obtener los siguientes resultados al trabajar con la aplicación DiLWO ([FR02b]) y en consultas varias:

- Encuentre todo los alumnos que están haciendo el ejercicio 2 del práctico 4
 - Ignacio Gómez en la sesión 3
 - Emilce Rojo en la sesión 3
 - Francisco Lagos en la sesión 2
- Deseo unirme a la sesión de Lagos.
- Buscar todos los alumnos que realizaron el diagrama 5 con un bucle.
- Compare el tiempos de los resultados finales de las sesiones 2 y 3
 - Sesión 2: 30 minutos
 - Sesión 3: 10 minutos

→ Busque todos documentos donde mencione la serie de Fibonacci.

— Se ha encontrado en el libro "Análisis 1".

— Se ha encontrado en el apunte "Resolución de Problemas y Algoritmos".

→ Si Claudia Fernández está trabajando en alguna sesión mandarle el siguiente mensaje: "Hiciste las llaves del 5?"

Existen herramientas en el mercado, tal como TextAnalyst [<http://www.megaputer.com>], que permite analizar un texto en detalle dando como resultado una lista de los conceptos más importantes, estableciéndolos luego como hipervínculos en el texto original. De esta manera se provee un mecanismo eficiente de navegación a través de texto, y se automatiza la creación de resúmenes, se clasifica el texto, se compara y se permite recuperar información utilizando el lenguaje natural. Además, se permite el manejo de grandes volúmenes de información, por lo tanto es una gran ventaja para el trabajo con el material didáctico de la materia y el resultante de la investigación de alumnos y profesores. Este material no se limita a documentos académicos sino también a correos electrónicos, páginas web, entre los más importantes.

En el caso del análisis de e-mails, estos se pueden:

- clasificar automáticamente de acuerdo al contenido, grupos de alumnos, etc.
- extraer mensajes donde se detecten dudas acerca de algún contenido.
- dar prioridad a determinados usuarios o grupos de personas.

Esto implica una disminución de la cantidad de tiempo requerido para el manejo y análisis de la información y simplifica la tarea de "Help Desk" o ayuda que el sistema pueda brindar a los usuarios.

4 Conclusiones

El procesamiento en lenguaje natural es un área relativamente nueva, las investigaciones están orientadas a conseguir mejoras sustanciales en diversas áreas de análisis tales como: la resolución del sentido de la palabra; el desarrollo de técnicas robustas que permitan la identificación del idioma, reconocer y clasificar entidades y resolver anáforas o correferencias; el desarrollo de sistemas que sean capaces de agrupar e interrelacionar los documentos, generar automáticamente resúmenes de documentos o grupos de documentos, producir una representación conceptual de consultas en cualquier idioma deseado.

Tanto alumnos como docentes, cuentan con un caudal cada vez mayor de información, lo cual se ve incrementado con el uso de la web y avances tecnológicos. El procesamiento en lenguaje natural brinda un mecanismo que facilita el acceso a dicha información reduciendo la complejidad de la recuperación de la información.

Las herramientas, disponibles en el mercado actualmente, son rudimentarias. Sin embargo, a medida que los investigadores sigan avanzando, contaremos con un potencial cuantioso en calidad que nos permitirá brindar una mejor calidad en la educación a distancia.

La plataforma presentada en el trabajo [FR02b] brinda un marco propicio para el desarrollo del dictado de la materia RPA a distancia. Sin embargo, dado que el volumen de información al que acceden docentes y alumnos, crece exponencialmente, es necesario recurrir a nuevas herramientas que agilicen la recuperación, análisis y comunicación de la misma.

Referencias

- [AHM00] Xavier Alami n, Pablo Haya, and Germ n Montoro. Odisea: Hacia un entorno inteligente basado en un interfaz en lenguaje natural. *I Jornadas de Tecnolog a del Habla. Sevilla. Espa a*, 6 al 10 de Noviembre. 2000.
- [All95] James Allen. *Natural Language Understanding*. The benjamin/Cummings Publishing Company, Inc. 2nd ed., 1995.
- [FR02a] Carina Fracchia and Jorge Rodriguez. Plataformas de educaci n a distancia: Consideraciones para su implementaci n. *VIII Congreso Argentino en Ciencias de la Computaci n.*, 2002.
- [FR02b] Carina Fracchia and Sandra Roger. Propuesta para el dictado a distancia de la materia resoluci n de problemas y algoritmos. *VIII Congreso Argentino en Ciencias de la Computaci n.*, 2002.
- [Mit01] Ruslan Mitkov. Outstanding issues in anaphora resolution. *Second International Conference, CICLing 2001*, pages 110–125, 2001.
- [Ped01] Ted Pedersen. Lexical semantic ambiguity resolution with bigram-based decision trees. *Second International Conference, CICLing 2001*, pages 157–168, 2001.
- [PSNn⁺01] Manuel Palomar, Maximiliano Saiz-Noeda, Rafael Mu noz, Armando Suarez, Patricio Mart n e z-Barco, and Andr s Montoyo. Phora: A nlp system for spanish. *Second International Conference, CICLing 2001*, pages 126–139, 2001.
- [Rog01] Sandra Roger. Belief revision on anaphora resolution. *Second International Conference, CICLing 2001*, pages 140–142, 2001.