Un ed, radioeducativa. Hola, buenas tardes. Reciban un saludo de Rosa Sánchez desde el control y de Miguel Minaya ante el micrófono en nombre del equipo que realiza este programa desde la Universidad Nacional de Educación a Distancia. Comienza ahora un tiempo de radio en el que se va a hablar de la educación a distancia. Hola, buenas tardes. Reciban un saludo de Rosa Sánchez desde el control y de Miguel Minaya ante el micrófono en nombre del equipo que realiza este programa desde la Universidad Nacional de Educación a Distancia.

el que la educación y la cultura son protagonistas. Nos sintonizan a través de Radio 3, de Radio Nacional de España en FM y en el 1359 de la Onda Media. También pueden escucharnos a través de Internet, acudiendo a la página web de nuestra universidad, uned.es. Eso sí, recuerden que en Internet encontrarán nuestro programa pasada una semana de la fecha de emisión en radio. Cada fin de semana la UNED les ofrece 10 horas de programas educativos que se realizan íntegramente desde el CEMAP, Centro de Medios Audiovisuales, en un intento de acercar la universidad a todos nuestros alumnos y oyentes en general. Hoy sábado nuestros espacios están relacionados con buena parte de las carreras y cursos que pueden seguirse a través de la UNED. En la primera parte de nuestra emisión nos vamos a dedicar a las facultades de Filología, Geografía y Ciencia.

Historia y Ciencias, además del espacio La Voz de la Bizarra, del programa de formación del profesorado. Pasado este tiempo volveremos a dar un avance del resto de la emisión. Bien, y ya vamos a entrar de lleno en el contenido educativo de la emisión de hoy, 25 de marzo, dando paso a nuestro primer espacio. Se trata de Aula de Filología. En Aula de Filología de esta noche nos acompañan Tim Ritt. Tim, buenas tardes. Hola, buenas tardes a todos. Y Elena Bárcena. Elena, buenas tardes. Buenas tardes, Miguel. Tim Ritt es profesor de la Escuela de Ingeniería en la UNED. Y Elena Bárcena es profesora en la Facultad de Filología de la UNED también. El título que figura en la guía es El Paseo de la Filología de la Universidad de Madrid. El Paseo de la Filología de la Universidad de Madrid.

El papel del procesamiento del lenguaje natural en la web. Tim, antes de hablar sobre el procesamiento del lenguaje natural en la web, podrías hablar un poco sobre la historia de la web para hacer un poco de introducción, ¿no? Centrados a los alumnos o a los oyentes que nos escuchen. Pues adelante. Y antes de comentar lo que es la base tecnológica de la web y un poco de la historia, lo que quería comentar, más que nada recomendar, es un buen libro. Porque para los estudiantes que les interesa mucho el tema, y francamente la web está en todas las partes hoy en día, está en la televisión, en la radio, en los periódicos, sale en las direcciones web, en muchos sitios, para mucha gente les da un poco de miedo. Entonces, antes de empezar con el tema de hoy, quería comentar que lo que mi experiencia es un poco... Es un buen libro. Y se llama ¿Cómo funciona Internet? Es un libro escrito por Preston Grala con el editorial Prentice Hall. Y como he comentado antes, yo creo que merece la pena. Se trata de un libro con una base tecnológica.

muy básico. Entonces, no hay que tener miedo o temar el tema. Bien. Entonces, vamos a empezar. La web. Viene, en primer lugar, la palabra. Viene del inglés World Wide Web, o lo que se suele llamar en varios sitios el telearaña mundial. Y funciona. Es una tecnología relativamente nueva. Y funciona sobre una base que se llama la Internet, que es una base mucho más antigua. Muy bien. Bien.

de defensa del gobierno de los Estados Unidos, y luego fue siendo adoptado por universidades y centros de investigación como vía de comunicación interna entre los distintos centros. Funciona sobre el protocolo de intercambio de paquetes electrónicos de información, que se llama Internet Protocol, o quizás nuestros oyentes estén más familiarizados con el acrónimo IP. Y se trata de un modelo servidor-cliente, en el que la máquina servidor es una potente máquina central que contiene información y programas, y luego hay una serie de máquinas que tienen acceso a esta información, que son las máquinas clientes, pequeñas máquinas. Bien, entonces sobre la Internet, la web empezó como un proyecto de investigación en las redes sociales. Y luego...

El proyecto principal de los investigadores entonces era desarrollar una manera de acceder y compartir información sin tener conocimientos informáticos muy específicos, muy profesionales. Perdona Tim, creo que antes de seguir deberías comentar lo que es CERN, ¿no te parece? Sí, claro. CERN es el Centro Europeo de Investigaciones Nucleares. Y empezaron con la idea de hacerlo más accedible estos materiales. Entonces empezaron a desarrollar protocolos y herramientas para acceder a materiales principalmente textuales. Entonces empezaron con un lenguaje de etiquetación que se llama hipertexto, que consiste básicamente en textos, también con algunas imágenes. Y sonidos y lo que son enlaces, es decir, para mantener una estructura jerárquica de información para que una persona si está leyendo un texto encuentre una palabra justamente.

Como hemos encontrado aquí la palabra son y quieren saber lo que es, y si está conectado a una definición, puede poner el ratón, el puntero sobre ello y pinchar para llevarle a una definición de dicha palabra. Es una buena manera de organizar la información. Se accede a la información de Internet a través de un programa que se llama navegador. Seguramente casi todo el mundo está familiarizado con este término. Los principales son Internet Explorer de Microsoft o el Navigator de Netscape, pero también hay muchos otros. Y es interesante hacer hincapié en la noción de navegación, que es la búsqueda en un espacio virtual de direcciones IP, o sea, de Internet Protocol. También a veces se utiliza el término de surfear en la web, ¿no? Navegar o surfear en la web. Es muy demócrata.

con las jóvenes. Sí, hemos tratado en un programa anterior de radio lo que es el procesamiento del lenguaje natural. Diré el acrónimo, que es PLN, lo que pasa que viéndolo escrito como lo tengo yo, pues es más sencillo, pero en radio es más difícil de retener. ¿Cómo puede encajarse el PLN, el procesamiento del lenguaje natural, en la web, Elena? Pues es evidente que hace falta más que materiales textuales para que realmente podamos empezar a hablar de procesamiento del lenguaje natural en la web. Son necesarios programas en el cliente del usuario, la máquina cliente del usuario, y o también en el servidor, más que páginas de texto, como digo. No sé, quizás podríamos poner algunos ejemplos más adelante. Sí, y yo creo que antes de ver los ejemplos debemos aclarificar un poco la tecnología, la tecnología subyacente a estos programas.

Principalmente, un término que se suele oír mucho en este campo es lo de los scripts CGE. Es decir, que son programas colocados en el servidor que un usuario de la web puede acceder a través de páginas web. Entonces, un ejemplo muy sencillo sería lo de las páginas de búsqueda que están en la web. Por ejemplo, Yahoo. Eso es lo que es la tecnología base. Pero más adelante vino la tecnología Java, que es más que un simple lenguaje de programación, que por cierto, cuyo aspecto es parecido a C o C++, si hay alguien

escuchando que les interese este tema. Pero, como he dicho, es más que un simple lenguaje, porque nos permite una manera de preparar programas que funcionen de una manera independiente de la plataforma en que están funcionando. ¿Qué quiero decir? Lo que acabo de decir.

que en vez de escribir un programa que solo va a funcionar, por ejemplo, en una máquina de Windows o un servidor Unix, podemos preparar los programas para que corran en todas las máquinas. Y esto lo hacemos a través de una máquina virtual que es simulando un entorno para este lenguaje. Y los resultados son bastante espectaculares porque además de tener la posibilidad de procesar textos metidos en una página web, también podemos tener simulaciones metidas en una página web, es decir, animaciones de gráficos y programas con interfaces bastante sofisticados. Y hay muchos programas hoy en día en la web que son pequeñas applets. Applets vienen del término en inglés pequeña aplicación, applet, aplicación. Y más adelante veremos algunos ejemplos. ¿Debería? También decir que hay problemas con el tiempo necesario para...

cargar los applets. De hecho, irónicamente a veces uno se refiere a la World Wide Web como World Wide Wait, que es esperar. Es un problema que no es trivial porque si hay que esperar 5, 10 o incluso más minutos para acceder a una información, pues el usuario acabará por intentar acceder a la información que necesita de otro modo y simplemente abandonar la web. Pero bueno, para ver realmente las posibilidades de estas tecnologías, quizá el mejor modo sería verlo a través de ejemplos. Sí, pues yo siempre recurro, Elena, a los ejemplos porque creo que es lo más ilustrativo y sería quizás ahora un buen momento para ver algunos ejemplos de programas disponibles en la web que conllevan algún aspecto del PLN. No me resisto a decir solo el acrónimo, tendría que decir el examen del lenguaje natural para familiarizarnos, ¿no? Sí, sí. Bien, podemos hablar, perdón. Sí. ¿Qué te parece de 5 sistemas?

Por ejemplo, el análisis de corpus con TACT Web. Se trata de la versión en la web de un programa muy conocido, la verdad que en el entorno del procesamiento del lenguaje natural, que ha sido pionero en su especialidad, que es el análisis, sobre todo, cuantitativo de Córpora. Pero bueno, también existe una versión como una aplicación normal y corriente que irónicamente también se puede bajar de la web. Bien, otro sistema del que podremos hablar será la traducción automática con sistemas como Sistran, que está en la página web de Alta Vista y se puede acceder desde allí. También hay un psiquiatra virtual llamado Elisa, que en realidad es el primer sistema que hubo de inteligencia artificial aplicada al procesamiento del lenguaje natural. Y aquí también se puede acceder al proceso de realización de la traducción del lenguaje natural, que es el proceso de bible berries, que es el proceso de matadero. Y aquí también se puede acceder a las mercancías unificadas para recobrar la sed de las bombas. Y aquí también se puede acceder a la red de mataderas que se aplican para crear las languajes, donde se otorga una aplicación automática autónoma de los magmáticos. La edad de girar un mandatario es en la serie de datos Development Which Swaps o de不知道 ahead s. Es una herramienta original de Googleious donde se reconnace fundamentalmente el valor importancia de unayondaspring a disconnect立 parte de la comunidad.

Y, bueno, la verdad es que es un caso muy curioso y anecdótico, aunque la verdad es que… Hay un buen anecdoto relacionado con el tema que la persona que lo hizo,

Faisenbaum, que lo hizo en 1966, cuando terminó el programa, llamó a su secretaria y que vino al despacho para dejar de jugar un poco con el programa y después, me parece que después de como cinco minutos, ella pidió a él que, por favor, que salga de la habitación, porque estaba un poco íntima esta habitación. Se sentía un poco intimidada. Y eso prueba hasta qué punto es efectivo. Sí, la verdad es que es un caso muy simpático de… Curioso, desde luego, sí. De aplicación, sí. Luego también está el sistema de ayuda GIFS, que, bueno, consiste en… Bueno, se hacen una serie de preguntas para obtener acceso a información de la web a través de enlaces. Y luego, por último, si nos da tiempo… Hablaremos de un sistema de enseñanza…

de inglés que hemos desarrollado Tim y yo, llamado el UPV. Y, bueno, también como otro ejemplo de procesamiento del lenguaje natural en la web. Muy bien. Nos parece que pasemos a desglosar en todos los ejemplos. Tim, ¿qué es TagWeb? Bien. Yo creo que ya hemos hablado un poco sobre las herramientas de análisis de corpos en un programa anterior. Entonces, no voy a volver sobre el tema. Pero, como comentó Elena antes, lo que nos da TagWeb es la posibilidad de trabajar con una herramienta de este tipo a través de una página web. Entonces, más esta página web y más adelante Elena dará la dirección y se encontrará con varias cajitas de texto en que un usuario puede hacer una pregunta. Me parece que los textos que están en la web son obras de Shakespeare para evitar problemas de los derechos de autor. Entonces, efectivamente, podemos buscar palabras y buscar frases y trabajar con los textos.

hace análisis estatísticos y funciona muy bien. Y además, me parece, no sé, no me acuerdo exactamente qué universidad, Elena, quizás acuerdas tú, pero no sé si es una competencia, pero por lo menos en una universidad por aquí hay unos profesores que están usando esta aplicación para algunas de sus sesiones de laboratorio en sus clases de procesamiento de lenguaje natural. Y es una manera muy efectiva de trabajar con este tipo de aplicación, porque además de estar en el laboratorio, un usuario puede volver a trabajar con ello desde su propia casa. Sí, el tipo de información que podemos obtener, contactos, por ejemplo, de ocurrencia de palabras, o de conjuntos, o de secuencias de palabras. Siempre solemos trabajar a nivel de cadena lingüística, a nivel superficial. Entonces, por ejemplo, cuántas veces aparece en el texto... Se me ocurrió algo.

Ahora, recordando los días en que estudiaba literatura inglesa, es muy importante el concepto de oído en Hamlet, en Shakespeare, por aquello del modo en que es asesinado el padre de Hamlet, con el veneno en el oído. Entonces, hay unas referencias a la palabra oído a lo largo de todo el texto de Hamlet. Entonces, bueno, pues podíamos preguntar automáticamente, en vez de tenernos que leer y escanear manualmente todo el libro de Hamlet, toda la obra, pues preguntar simplemente a Tag Web cuántas veces y en cuáles son las instancias, en qué puntos aparece la palabra oído. Y me parece que tienes la dirección web ahí, Elena. Sí, la voy a dar a continuación. Es la siguiente. Pigeon, en inglés, o sea, P-I-G-E. P-I-G-E.

.uk barra tactweb t-a-c-t-w-e-b barra doc d-o-c barra tact.htm t-a-c-t .htm Ya no hace falta aquello del http dos puntos doble barra porque ahora ya lo ya lo saben todos los navegadores y tenemos las mismas funciones en el programa de la web que en el programa normal por si alguien está familiarizado con el programa de la Universidad de Toronto original que elaboró John Bradley hace muchísimos años pues los mismos análisis

estadísticos gráficos el programa de QUIC que es Keyword in Context el acrónimo o sea que si alguien está

familiarizado con el programa normal, pues es similar lo que encontrará en la web. Sí, y las ventajas que tiene la versión en la web, claro, es que cualquier persona con un navegador normal y corriente puede usar el programa sin tener una versión propia en casa, que está muy bien. Perfecto. Pasamos ahora a comprobar el segundo ejemplo que comentaste, Elena, Sistram, ¿verdad? Pues con Sistram sucede un poco como contacto, que tenemos la versión en la web y mucha gente seguramente conoce Altavista, que es una máquina de búsqueda en la web. Bueno, pues allí podemos encontrarlo. Una máquina de búsqueda, bueno, es un poco, supongo que lo que ha dicho antes Tim, se teclea una palabra y entonces se buscan las palabras relacionadas que existen en su base de datos en Internet, ¿no? Claro, y la ventaja de tener la traducción automática en una página web es que aumenta las posibilidades de que la persona pueda usar la versión en la web. Bueno, pues aquí tenemos la versión en la web.

posibilidades que vamos a encontrar temas relacionados a lo que queremos saber. Si vamos a entender las páginas que encontremos, no sé, pero... La dirección es www.altavista.com Y vamos ahora a hacer una práctica. ¿Cómo funciona entonces, Tim? Bien. Abajo en la página de Altavista hay un enlace a una página en la que se encuentra el programa de traducción. La página consiste en una caja de texto y un menú plegable para elegir el par de idiomas y un botón para pasar el texto al programa de traducción. Entonces, Sistran hará la traducción y devolverá el resultado a la caja de textos. Y yo creo que la mejor manera de evaluarlo es ver un pequeño conjunto de ejemplos. ¿Sí te parece, Elena? Sí, perfectamente. Yo diré la versión en inglés y Elena, tú dices lo que produce el programa. Muy bien. Adelante. Bien. The child had a ball. Pues obtenemos con el sistema...

niño tenía una bola, que quizá no es la mejor traducción. Supongo que, no sé, no sé por qué precisamente han elegido esta traducción los investigadores que han desarrollado Sistran, porque quizá la traducción inmediata que uno piensa de ball en español es pelota, ¿no? Pero, bueno, puede ser porque en el sublenguaje correspondiente administrativo hay alguna circunstancia, algún contexto en el que ball encuentre la traducción como bola. Y seguramente también indica que el programa está trabajando con reglas, que están intercambiando palabras y no está tomando tanto en cuenta el contexto, porque debido al conocimiento del mundo que tenemos, cuando se trata de un niño con una pelota, sabemos que es una pelota y más bien un baile o cualquier otro tipo. Exactamente.

Lo demás está bien traducido, es decir, la concordancia morfológica entre el artículo y el sustantivo, el niño, también la morfología del verbo, su tiempo en pretérito imperfecto, el niño tenía, está todo correcto. Pero quizás sí, si acaso el único comentario negativo es precisamente el haber establecido la relación semántica o del mundo real en la que un niño se relaciona principalmente con una pelota, aunque no siempre. Vaya, también no tenemos suficiente contexto porque esto es una oración aislada, habría que ver dentro de un texto, quizás también podría haber un caso en el que ball se tradujera como bola. Bien, vamos a ver un ejemplo. Por ejemplo, tomorrow night I am going to a ball. Voy mañana por la noche a una bola. Una cosa que quizás sorprenda a nuestros oyentes...

cuya lengua materna seguramente será el español, es que aunque el orden de los elementos en la oración es muy flexible en español, bueno, pues quizás sería más natural en español poner el circunstancial de tiempo mañana por la noche, bien al principio de la oración o al final. Mañana por la noche voy a una, bueno, a lo que sea, o voy a no sé dónde mañana por la noche, y no quizá en medio de la oración. Y aquí nos damos cuenta también al ver la traducción de ball por bola, que no se ha tenido en cuenta ningún tipo de información semántica o contextual, porque ball tendría que haberse traducido como baile, así que seguramente este es un simple defecto del vocabulario, el diccionario bilingüe, el lexicón bilingüe del sistema. Y bueno, pues simplemente esto se podría haber solucionado con un pequeño simbólico. Me ocurre así, de pronto con un pequeño...

rasgo semántico que indique que cuando volva acompañado de una referencia temporal o del verbo ir, en fin este tipo de información pues entonces debería traducirse como baile Muy bien El siguiente ejemplo que comentó Elena al principio es el del psiquiatra virtual Elisa, que yo tengo que decir que me ha dejado un dato sorprendido Sí, es muy curioso Lo que suele pasar con Elisa es que la gente empieza con muchas ganas, pero después de como no sé, cinco o diez minutos se dan cuenta que está repitiendo un poco y no dando mucha información Sí, pero hay que reconocer que para ser un programa que se creó en 1966 es impresionante que tardemos cinco o diez minutos en darnos cuenta de que no hay un hombrecito dentro del ordenador hablando con nosotros Bien, entonces yo creo que se trata básicamente de un sistema de diálogo en lo cual el sistema analítico full handle su dialito puede con trozos al hogar el dominante que tiene si el buscan el paso junto a la dentist si laомина al centro de acción en el control on par Republican o en las carguerías de area pessimately y asíth porist al concepto Luis el compromiso mediante una mira la unión deálogo para transformar ¿انos Saltinella? Lasarse de Acasio Price nodes yo gardenios tanto, eso todo, etc.

lo que escribe el usuario y basado en unas reglas lingüísticas contesta. Y no vamos en este caso a dar ninguna dirección web porque es tan famosa simplemente entrando el nombre Eliza en cualquier máquina de búsqueda el usuario encontrará muchas direcciones. Pero yo creo que lo más divertido quizás sería que pasamos por un diálogo muy corto. Entonces yo voy a jugar el papel del usuario y Elena del sistema, si os parece. Bien. Estoy un poco preocupado. ¿Cuánto tiempo lleva un poco preocupado? No sé. Quizás estoy un poco ansioso. ¿Cree que es normal estar un poco ansioso? No sé. Mucha gente se preocupa. Quizás sea normal. ¿Qué le sugiere eso? Etcétera, etcétera. Y podemos continuar mucho tiempo. Y de todas formas animaríamos mucho a los oyentes que van a una página y prueben el sistema. Sí, pero no acabo de saciar mi curiosidad. ¿Cómo puede un programa producir estas respuestas? Bien. Pues en primer lugar el programa tiene que tener un sistema

Tendrá un parser, es decir, un analizador, que intentará acoplar partes del texto del usuario con reglas y entradas léxicas que tiene para poder formar las respuestas. Y algunos ejemplos serían, por ejemplo, la sustitución de las formas de los verbos, es decir, es por soy, está, estoy, etc. Y también la introducción de palabras claves, como por ejemplo, madre y padre, en el contexto de otras palabras, como por ejemplo, familia. Y por fin, si el programa no puede hacer ningún análisis de lo que ha entrado el usuario, siempre puede usar esas frases al azar, si no ha podido reconocer nada en la entrada. Sí, volviendo al texto, podríamos ver las primeras entradas una segunda vez, si os parece. Comenzaba Tim

diciendo, estoy un poco preocupado. Entonces, ¿esto será reconocido por el AISA como estoy? Más, bueno, pues...

un adjetivo, un participio en este caso. Entonces, una respuesta coherente sería, ya que además tenemos la forma de to be, que es estar, o sea, el equivalente de to be que tiene un carácter transitorio, bueno, pues con una pregunta de tipo temporal, ¿cuánto tiempo lleva preocupado, o podía haber sido alegre, o cualquier otro adjetivo, participio que el usuario hubiera introducido? Así que el sistema ignora completamente preocupado, o sea, que no sabe absolutamente nada de psicología, ni de terapia psicológica, ni nada por el estilo, sino simplemente reconoce el patrón sintáctico. Después de no haber reconocido nada, como la oración, no sé, quizás estoy un poco ansioso, entonces, tendría que responder con una frase al azar, y una frase así muy segura que...

Yo creo que se podría introducir, tras haber dicho Tim casi cualquier cosa, ¿cree que es normal estar un poco ansioso? A si pues Tim podría haber dicho, pues no sé, quizás me voy a ir ahora y voy a dejarte. ¿Crees que es normal hacerlo? Pero quizás después de haberlo oído unas cuantas veces, recibido el sistema, mejor dicho, pues nos chocaría. Pero claro, la primera vez pasa desapercibido. Sí, porque siempre podemos darnos cuenta que el sistema suele andar con respuestas abiertas y claro, a la hora de buscar cosas concretas, veamos las fallas del sistema. Brevemente, porque andamos mal de tiempo, el siguiente programa era el GIFS. ¿Qué es este programa? Bien, GIFS es un sistema de ayuda inteligente, es decir, que hace un análisis de preguntas que hacemos en lenguaje natural para poder presentar enlaces a las respuestas más apropiadas.

AskJeeves.com Funciona a través de, como decía Tim, del acoplamiento de partes de preguntas con partes de otras preguntas que tiene el programa almacenado en su base de datos. Y si te parece, Miguel, podríamos acabar con un ejemplo. Pues sí, perfecto. Sería lo mejor. Bueno, entonces, metiendo la pregunta ¿Cómo puede hacer mi rico? El programa produce un conjunto de preguntas que cree que son parecidos a lo que quiero saber. Entonces, por ejemplo, salió una lista ¿Cómo puedo hacer dinero? ¿Cómo puedo hacer mi rico rápidamente? ¿Y cómo puedo casarme para acceder al dinero? Evidentemente no hay ningún tipo de ética en este programa. ¿Y quiénes son los...? Hay unas categorías de famosos, actores, deportistas, etc. con más dinero del mundo. Y con cada pregunta hay un conjunto de enlaces. Es interesante una manera de acceder. ¿Al dinero?

Bien, pues no nos queda más tiempo es una pena porque imagino que esto daría mucho que hablar y bueno, los ejemplos además son los más sustanciosos, ¿verdad? Solo me queda despedirme y bueno, agradecer a Tim Ripp, profesor de la Escuela de Ingeniería de la UNED y Elena Bárcena de la Facultad de Filología de la UNED también su presencia en este espacio, en esta serie de espacios porque hoy era el tercero y último y ha sido una experiencia muy agradable y espero que los alumnos y los oyentes hayan seguido con atención sus propuestas Muchas gracias Tim, muchas gracias Elena y buenas tardes. Gracias a ti Miguel Gracias a todos Por nuestra parte nada más, termina aquí Aula de Filología. El próximo sábado, 1 de abril volveremos con un programa sobre literatura española. El profesor Vicente Granados traerá hasta nuestros estudios a la escritora Clara Sánchez, último premio Alfaguara con su novela Últimas Noticias del Paraíso Gracias por la atención que nos han prestado

Recuerden que la UNED les acompaña hasta las 12 de la noche. Ahora tienen ocasión de escuchar Revista de Geografía e Historia. Les dejo con nuestra compañera Pilar Carrasco.