

Corpus MAVIR - Acuerdo de licencia

Licencia expedida por el consorcio MAVIR, formado por los siguientes grupos de investigación de Madrid:

- Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid
- Grupo de Sistemas Inteligentes (GSI-ISYS) de la Universidad Politécnica de Madrid
- E.T.S.I. Informática de la Universidad Nacional de Educación a Distancia
- Grupo de Bases de Datos Avanzadas, Departamento de Informática de la Universidad Carlos III de Madrid
- Centro de Información y Documentación Científica (CINDOC-CSIC)
- Grupo de Sistemas Inteligentes de la Universidad Europea de Madrid

El usuario:

acuerda lo siguiente:

1. Definición

El corpus MAVIR es una colección de grabaciones de sonido y vídeo con sus correspondientes transcripciones de habla oral, procesadas informáticamente. Su elaboración se dirige a la investigación en procesamiento de lenguaje natural y tecnologías de habla.

2. Estatuto del titular de la licencia

El titular de la licencia confirma que pertenece a una institución académica, investigadora o educativa.

3. Contenidos del producto

El corpus MAVIR consta de trece grabaciones de habla oral espontánea (sonido y vídeo) en español y en inglés, que se realizaron durante las siguientes jornadas del consorcio MAVIR:

- I jornadas MAVIR, celebradas en la Universidad Politécnica de Madrid y la Universidad Nacional de Educación a Distancia (UNED), los días 16 y 17 de noviembre de 2006.
- II jornadas MAVIR, celebradas en la Universidad Europea de Madrid, los días 15 y 16 de noviembre de 2007.
- III jornadas MAVIR, celebradas en el Centro de Ciencias Humanas y Sociales (CSIC) de Madrid, los días 27 y 28 de noviembre de 2008.

Las conferencias y mesas redondas grabadas se distribuyen en archivos del siguiente modo:

- 1) mavir01: incluye la grabación y la transcripción de la conferencia de Ralph Grishman (New York University) "Challenges for Information Extraction", celebrada en el Salón de Actos de la Facultad de Económicas de la UNED el jueves 16 de noviembre de 2006.
- 2) mavir02: incluye la grabación y la transcripción de la mesa redonda (celebrada en el Salón de Actos de la Facultad de Económicas de la UNED el jueves 16 de noviembre de 2006) "Proceso de innovación de tecnologías de acceso a la información: ¿Cómo llegar al mercado?", moderada por Anselmo Peñas (UNED), en la que participaron las siguientes empresas: Bitext, Daedalus, iSOCO, Thera-CLIC y Eleka.
- 3) mavir03: incluye la grabación y la transcripción de la conferencia de Adolfo Corujo "España y los buscadores: un mercado potencial", celebrada en la Sala Verde (1ª planta) de la ETSI Caminos (UPM)
- 4) mavir04: incluye la grabación y la transcripción de la presentación "Aplicaciones de Tecnologías de la Lengua en Dominios médico y cultural", celebrada en la Sala Verde (1ª planta) de la ETSI Caminos (UPM): "Extracción de Información de los historiales médicos" (Manuel de Buenaga, UEM); "Construcción de rankings de universidades de acuerdo a indicadores Web" (Isidro Aguillo, CINDOC-CSIC); y "Gestión de flujo documental en los historiales médicos" (Paloma Martínez Fernández, UC3M)
- 5) mavir05: incluye la grabación y la transcripción de la conferencia de Satoshi Sekine (profesor/investigador asociado de la New York University) "On-demand Information Extraction", celebrada el jueves 15 de noviembre de 2007.
- 6) mavir06: incluye la grabación y la transcripción de la presentación "Buscador General Panhispánico", celebrada el 16 de noviembre de 2007, a cargo de Rosa Arandilla (Instituto Cervantes), Luis A. Camarero (Telefónica Soluciones) y Julio Villena (Daedalus).
- 7) mavir07: incluye la grabación y la transcripción de la presentación "Tecnología de la Web Semántica en Atención al Cliente", a cargo de Jesús Contreras (iSOCO) y Fermín Lozano (11811 Nueva Información Telefónica) el viernes 16 de noviembre de 2007.
- 8) mavir08: incluye la grabación y la transcripción de la presentación del ganador del I Premio MAVIR, a cargo de Francisco M. Rangel, el viernes 16 de noviembre de 2007.
- 9) mavir09: incluye la grabación y la transcripción de la presentación "Buenas prácticas en presencia web para grupos de investigación", a cargo de Isidro Aguillo, celebrada el 12 de noviembre de 2007 en el salón de actos de la Facultad de Psicología (UNED)
- 10) mavir10: incluye la grabación y la transcripción de la conferencia de Paul Clough (profesor de la University of Sheffield) "Multimedia Retrieval and Evaluation", celebrada el jueves 27 de noviembre de 2008.
- 11) mavir11: incluye la grabación y la transcripción de la presentación del premio MAVIR 2008, a cargo de Álvaro Rodrigo Yuste, celebrada el jueves 27 de noviembre de 2008.
- 12) mavir13: incluye la grabación y la transcripción de la presentación "Buscando cangrejos en Flickr", a cargo de Julio Gonzalo (UNED), celebrada el 28 de noviembre de 2008.

A las personas grabadas en el corpus MAVIR se les pidió permiso por escrito, y firmaron su consentimiento para la realización del presente producto a partir de sus grabaciones.

La información ofrecida es la siguiente:

- Convenciones de transcripción.
- Grabaciones de sonido (formato WAV).
- Grabaciones de vídeo (formato MPG2).
- Transcripciones de habla (presentadas en formato XML y TXT).

El consorcio que expide la licencia y el titular de la licencia reconocen que los derechos de copia (copyright) de los contenidos de las conferencias celebradas en las jornadas MAVIR pertenecen a los respectivos conferenciantes.

Todos los derechos, incluyendo los derechos de copia (copyright), que no se garantizan al titular de la licencia en este contrato, son propiedad de las instituciones que participan en el consorcio MAVIR, indicadas al inicio del documento.

4. Créditos y copyright de los contenidos

El usuario y la institución que expide la licencia reconocen los derechos de autor de los respectivos participantes en las conferencias y mesas redondas MAVIR:

- 1) Ralph Grishman (New York University), la conferencia “Challenges for Information Extraction”.
- 2) Los participantes en la mesa redonda “Proceso de innovación de tecnologías de acceso a la información: ¿Cómo llegar al mercado?”:
 - Enric Castellón (Thera-CLIC).
 - Iñaki Irazabalbeitia Fernández (Eleka).
 - José Luis Martínez Fernández (Daedalus).
 - Anselmo Peñas (chairperson, UNED).
 - Luis Rodrigo Aguado (iSOCO).
 - Antonio Sánchez Valderrábanos (Bitext).
- 3) Adolfo Corujo (ALMA Technologies), la presentación “España y los buscadores: un mercado potencial”.
- 4) Los participantes en la presentación de los proyectos de investigación en aplicaciones tecnológicas del habla en el dominio médico y cultural:
 - Manuel de Buenaga (UEM): “Extracción de Información de los historiales médicos”.
 - Isidro Aguillo (CINDOC-CSIC): “Construcción de rankings de universidades de acuerdo a indicadores Web”.
 - Paloma Martínez Fernández (UC3M): “Gestión de flujo documental en los historiales médicos”.

- 5) Satoshi Sekine (New York University), la conferencia "On-demand Information Extraction.
- 6) Rosa Arandilla (Instituto Cervantes), Luis Ángel Camarero (Telefónica Soluciones) y Julio Villena (Daedalus), la presentación "Buscador General Panhispánico".
- 7) Jesús Contreras (iSOCO) y Fermín Lozano (11811 Nueva Información Telefónica), la presentación "Tecnología de la Web Semántica en Atención al Cliente".
- 8) Francisco M. Rangel, la presentación "Premio MAVIR 2007".
- 9) Isidro Aguillo (CSIC-CINDOC), la presentación "Buenas prácticas en presencia Web para grupos de investigación".
- 10) Paul Clough (University of Sheffield): la conferencia "Multimedia Retrieval and Evaluation".
- 11) Álvaro Rodrigo Yuste (NLP group, UNED): la presentación del premio MAVIR 2008.
- 12) Marcin Detyniecki (Centre National de la Recherche Scientifique): la presentación "Beyond Text-based Multimedia Retrieval".
- 13) Julio Gonzalo (NLP group, UNED): la presentación "Buscando cangrejos en Flickr".

Además, el usuario reconoce que el corpus fue diseñado, transcrito y procesado informáticamente en el Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid por las siguientes personas:

Dirección: Antonio Moreno Sandoval

Transcripción y revisión:

- Sergio Calvo Páez
- Leonardo Campillos Llanos
- Marta Garrote Salazar
- Sara Garrote Salazar
- Ana González Ledesma
- Alicia González Martínez
- M^a. Cristina Tovar Pérez
- Ana Valverde Mateos

6. Licencia

El consorcio que expide la licencia garantiza una licencia no exclusiva para utilizar el corpus MAVIR.

El titular de la licencia acuerda:

- usar el corpus solo para objetivos de investigación no comerciales y no lucrativos;
- no realizar cambios en el corpus;
- y reconocer el uso del corpus en todas las publicaciones que aporten resultados producidos con la ayuda del corpus MAVIR.

Esta garantía de licencia está libre de pago por parte del consorcio que la expide.

El consorcio que expide la licencia la garantiza al titular de la misma libre de pago.

Usted acuerda que el titular de la licencia tiene el derecho de identificarle públicamente como usuario/titular de la licencia del corpus MAVIR.

Este acuerdo implica el reconocimiento de acuerdo en su totalidad entre las partes que conciernen al presente asunto, y no existen términos o condiciones expresos o implícitos escritos u orales del presente. Este acuerdo suplanta toda formalización, acuerdo, promesa u otra forma de comunicación anterior oral o escrita concerniente o relacionada al asunto de este acuerdo.

El uso del corpus o de los datos derivados del corpus para cualquier propósito comercial requiere acuerdo escrito explícito del consorcio que expide la licencia.

6. No revelación de datos y requisito de precaución profesional

El corpus MAVIR será mantenido confidencialmente por parte del titular de la licencia, quien no podrá copiar, publicar, distribuir o de cualquier otro modo ofrecer a terceras partes el corpus MAVIR, sea entero o en parte. El titular de la licencia asegurará que sus empleados comprenden y acatan esta restricción, y supervisará sus actividades con respecto al corpus.

El titular de la licencia acuerda mantener confidencial el corpus MAVIR y tratar tales materiales con al menos el mismo grado de cuidado que trata su propia información de mayor confidencialidad. El titular de la licencia no puede asignar, transferir, arrendar, vender, alquilar, cobrar o beneficiarse de otro modo a partir el corpus MAVIR.

7. Finalización de la licencia

Este acuerdo entrará en vigor desde la fecha de la firma y continuará hasta que finalice por parte del consorcio que expide la licencia o el titular de la misma.

El consorcio que expide la licencia puede finalizar el acuerdo inmediatamente si el titular de la misma realiza cualquier incumplimiento material de este acuerdo o si el dueño de los derechos de copia (copyright) plantea cualquier objeción a los términos del acuerdo de licencia.

El titular de la licencia puede finalizar este acuerdo en 30 días de aviso por escrito. En la finalización de la licencia de cualquier manera producida, el titular de la licencia borrará todas las copias del corpus MAVIR. En la finalización, el titular de la licencia borrará o destruirá de cualquier otro modo toda copia total o parcial del corpus MAVIR. Un agente debidamente autorizado del titular de la licencia verificará enseguida que el titular de la licencia ha cumplido con su obligación escribiendo al Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid.

8. Renuncia de responsabilidad

Esta licencia se garantiza con la premisa de que el dueño de los derechos de copia (copyright) de los datos crudos (datos fuente) acepta los términos del acuerdo de licencia. Si el dueño de los derechos de copia (copyright) plantea una objeción a los términos del acuerdo de licencia, el consorcio que expide la misma puede finalizar el acuerdo y en consecuencia el titular de la licencia borrará todas las copias del corpus MAVIR.

Ninguna responsabilidad se acepta por cualquier daño causado por el uso de este corpus. El consorcio que expide la licencia informa por la presente al titular de la misma que la versión distribuida del corpus MAVIR puede presentar errores de anotación.

TITULAR DE LA LICENCIA

Institución: _____

Nombre: _____

Tratamiento: _____

Firma:

Fecha: