

ÍNDICE

ÍNDICE	1
DISEÑO DEL RECOMENDADOR	1
PROCESAMIENTO DE CAMPOS	1
CAMPOS DE USUARIO	1
CAMPOS DE ÍTEM	2
MEDIDAS DE SIMILITUD	2
SIMILITUD DE USUARIOS	2
SIMILITUD DE ÍTEMS	3
METODO DE RECOMENDACION	3
ETAPA OFFLINE	3
ETAPA ONLINE	3

DISEÑO DEL RECOMENDADOR

El recomendador en cuestión tendrá como objetivo recomendar problemas de trigonometría (de ahora en más “ítems”) a profesores de matemática (de ahora en más “usuarios”).

Los datos sobre los que se diseñará serán obtenidos mediante una encuesta donde se recolectan problemas que suelen enseñarse en el nivel medio.

PROCESAMIENTO DE CAMPOS

A continuación se enumeran los campos de datos obtenidos de la encuesta, los cuales han sido clasificados según se relacionen al usuario que responde o al ítem a recomendar. Se detalla su contenido esperado de acuerdo a nuestras propias suposiciones y el procesamiento necesario sobre el mismo.

CAMPOS DE USUARIO

- **País donde se desarrolla como profesor de matemática:** Este campo contendrá el nombre de un país escrito en idioma español. No requiere ningún procesamiento adicional.
 - **Provincia donde se desarrolla como profesor de matemática:** La existencia de este campo es opcional y está ligada a que el valor del campo anterior sea “Argentina”. Contendrá el nombre de una provincia argentina o “null” si la entrada se refiriera a otro país. No requiere ningún procesamiento adicional.
- **Enseña o ha enseñado trigonometría:** Este campo contendrá la respuesta “sí” o “no” escrita en idioma español. Debe convertirse a booleano.
- **Curso de mayor experiencia como docente:** Este campo contendrá un curso basado en la edad de sus estudiantes. Debe convertirse a entero extrayendo el número de curso y eliminando la descripción de las edades.

- **Modalidad de su escuela con mayor carga horaria:** Contendrá la modalidad seleccionada de la lista de opciones contempladas en el formulario. En caso de que la respuesta provenga de la opción “Otro”, esta estará escrita en idioma español. No requiere ningún procesamiento adicional.
- **Orientación de su escuela con mayor carga horaria:** Contendrá la orientación seleccionada de la lista de opciones contempladas en el formulario. En caso de que la respuesta provenga de la opción “Otro”, esta estará escrita en idioma español. No requiere ningún procesamiento adicional.
- **Temas enseñados por año:** Este campo contendrá una respuesta escrita en idioma español. Debe dividirse el texto en tokens y eliminar las “stop-words”. Opcionalmente y según los resultados obtenidos durante la implementación puede aplicarse stemming y/o acotar el conjunto de tokens a las palabras más representativas.

CAMPOS DE ÍTEM

- **Curso en el que fue dado el problema:** Este campo contendrá un curso basado en la edad de sus estudiantes. Debe convertirse a entero extrayendo el número de curso y eliminando la descripción de las edades.
- **Conocimientos previos necesarios:** Este campo contendrá una respuesta escrita en idioma español. Debe dividirse el texto en tokens y eliminar las “stop-words”. Opcionalmente y según los resultados obtenidos durante la implementación puede aplicarse stemming y/o acotar el conjunto de tokens a las palabras más representativas.
- **Claves para su solución:** Este campo contendrá una respuesta escrita en idioma español. Debe dividirse el texto en tokens y eliminar las “stop-words”. Opcionalmente y según los resultados obtenidos durante la implementación puede aplicarse stemming y/o acotar el conjunto de tokens a las palabras más representativas.
- **Enunciado:** Este campo contendrá una respuesta escrita en idioma español. No requiere ningún procesamiento adicional.
 - **Imagen enunciado:** Este campo es opcional y puede contener la referencia a una imagen. No requiere ningún procesamiento adicional.
- **Solución:** Este campo contendrá una respuesta escrita en idioma español. No requiere ningún procesamiento adicional.

MEDIDAS DE SIMILITUD

A continuación se enumeran las medidas de similitud definidas para la comparación de usuarios e ítems.

SIMILITUD DE USUARIOS

La medida de similitud entre dos usuarios estará dada por la suma pesada de la comparación de cada uno de los siguientes campos:

- **País donde se desarrolla como profesor de matemática:** El resultado de la comparación será 1 si los campos son iguales o 0 en caso contrario.
 - **Provincia donde se desarrolla como profesor de matemática:** El resultado de la comparación será 1 si los campos son iguales o 0 en caso contrario. Si uno de los campos es “null” el resultado siempre será 0.

- **Enseña o ha enseñado trigonometría:** El resultado de la comparación será 1 si los campos son iguales o 0 en caso contrario.
- **Curso de mayor experiencia como docente:** El resultado estará dado por $1 - \frac{|x-y|}{6}$ siendo x e y los campos a comparar.
- **Modalidad de su escuela con mayor carga horaria:** El resultado de la comparación será 1 si los campos son iguales o 0 en caso contrario.
- **Orientación de su escuela con mayor carga horaria:** El resultado de la comparación será 1 si los campos son iguales o 0 en caso contrario.
- **Temas enseñados por año:** El resultado estará dado por la medida de similitud de Jaccard $\frac{|X \cap Y|}{|X \cup Y|}$ siendo X e Y los conjuntos de tokens a comparar.

SIMILITUD DE ÍTEMS

La medida de similitud entre dos ítems estará dada por la suma pesada de la comparación de cada uno de los siguientes campos:

- **Curso en el que fue dado el problema:** El resultado estará dado por $1 - \frac{|x-y|}{6}$ siendo x e y los campos a comparar.
- **Conocimientos previos necesarios:** El resultado estará dado por la medida de similitud de Jaccard $\frac{|X \cap Y|}{|X \cup Y|}$ siendo X e Y los conjuntos de tokens a comparar.
- **Claves para su solución:** El resultado estará dado por la medida de similitud de Jaccard $\frac{|X \cap Y|}{|X \cup Y|}$ siendo X e Y los conjuntos de tokens a comparar.

METODO DE RECOMENDACION

El recomendador a implementar contará con una etapa offline que tomará conceptos vistos en las técnicas de filtrado colaborativo usuario - usuario y que servirá como entrada a una etapa online con un enfoque más cercano a la recomendación basada en contenido.

ETAPA OFFLINE

En esta etapa se buscará la formación de clusters de ítems agrupados en base a la similitud de los usuarios que los cargaron, tomando como referencia el conjunto de atributos que consideramos que mejor describe a un usuario. Estos clusters multidimensionales nos permitirán obtener el conjunto de datos reducido y más cercano al usuario activo, lo que agilizará la operación durante la etapa online del recomendador.

ETAPA ONLINE

En esta etapa se utilizará la técnica Rocchio para obtener el conjunto de ítems recomendables. Inicialmente se obtendrá el cluster de ítems más cercano al usuario y se calculará un centroide o prototipo en base a los requerimientos especificados por el mismo (como podría ser el campo “Conocimientos previos necesarios”). Los ítems más cercanos a este centroide formarán parte del conjunto recomendable. Antes de presentar los ítems al usuario, estos se ordenarán según $1 - \frac{|x-y|}{6}$, siendo x e y los campos “Curso en el que fue dado el problema” y “Curso de mayor experiencia como docente” para cada ítem de forma que el usuario reciba los problemas cargados por profesores con mayor experiencia

en el curso que fueron dados. Como alternativa a este último ordenamiento consideramos la opción de incorporar el campo “años de experiencia” en el formulario de la encuesta para priorizar las recomendaciones de problemas aportados por profesores con mayor experiencia en un curso en particular.