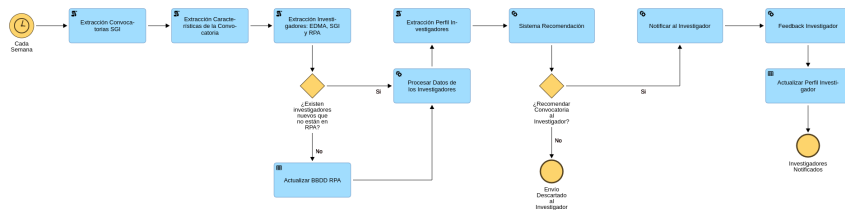


Proceso 4 : Sistemas Inteligentes de encaje entre convocatorias e investigadores



Descripción general del proceso.

Este proceso es el encargado de desarrollar un sistema de recomendaciones con el objetivo de poder ofrecer a los investigadores las convocatorias disponibles de la forma más personalizada posible.

Su funcionamiento comienza con la extracción de las convocatorias almacenadas en el SGI con estado REGISTRADA. Una vez obtenidas, se almacenará internamente toda la información con el objetivo de que los algoritmos de recomendación extraigan la información que deseen. Seguidamente, se extraerán todos los investigadores almacenados tanto internamente por la parte de RPA como en ED, en el caso de que hayan nuevos investigadores estos serán incorporados al sistema, se eliminan duplicados y los investigadores que hayan solicitado ya la convocatoria serán descartados como posibles candidatos a recomendar.

Los investigadores que se hayan insertado nuevos y no tengan un perfil configurado, el sistema tratará de cargar uno con la información en ED y SGI, en caso de que no fuera posible la carga de un perfil base por falta de información el investigador sería descartado, y el subproceso de "recordatorio de configuración de perfil" le notificará para que configure su perfil, de esta forma si el investigador establece su perfil la próxima ejecución el sistema de recomendación podría recomendarle la convocatoria que anteriormente no pudo.

Una vez extraídos los investigadores y procesados al igual que las convocatorias, esta información será introducida al sistema de recomendación el cual ejecutará los algoritmos que se describirán en detalle en las secciones posteriores, cada algoritmo tiene dos parámetros de entrada que el usuario funcional encargado del sistema debe introducir para programar la ejecución de este proceso, dichos parámetros son:

- **Activo.** Valor booleano que indica si ese algoritmo debe o no ejecutarse.
- **Peso.** Valor numérico cuya suma total de todos los pesos debe ser 1 e indica el peso que tiene el algoritmo a la hora de pasar los resultados al motor de recomendación (híbrido en este caso) para determinar si una convocatoria es susceptible de recomendar o no.

Una vez que los algoritmos terminen su ejecución, los datos de salida serán introducidos al motor híbrido (en la secciones posteriores se detallará su funcionamiento) que será el encargado de la decisión final. Aquellos investigadores que el sistema haya determinado que son aptos para recomendar recibirán un correo electrónico con la siguiente información de la convocatoria (pueden omitirse campos en caso de que no hayan sido introducidos en el SGI) :

- Entidad Convocante.
- Entidad Financiadora (podría haber más de una).
- Tipo Financiación.
- Porcentaje de Financiación.
- Importe de financiación.
- Título.
- Descripción.
- Observaciones.
- Fecha de publicación.
- ~~Fecha de concesión.~~
- Enlace a la convocatoria (SGI).
- Enlace para Feedback.
- Enlace para configurar el perfil.

En la siguiente imagen podemos ver un ejemplo del formato:

Hola Jose Manuel Bernabe Murcia,

En el SGI hay una convocatoria disponible, pensamos que te puede interesar.

Convocatoria

Entidad Convocante:EUROPEAN COMMISSION

Entidad Financiadora:EUROPEAN COMMISSION

Tipo Financiación:Subvención

Porcentaje de Financiación:100.0

Título: Prueba Título de la referencia CGF/OPE-02

Descripción: prueba de csp menus

Observaciones:

Ámbito Geográfico: Europeo

Fecha publicación: 14/02/2021

Fecha de concesión: 16/12/2021

Consulta la convocatoria en el SGI para más detalles

[Enlace a la convocatoria](#)

****¿Te ha resultado de interés esta convocatoria? [Si](#) / [No](#)**

Puedes obtener convocatorias más personalizadas configurando su perfil.

[Configurar mi perfil](#)

****No comparta ningún enlace, son exclusivos para cada investigador.****

© Hercules-RPA

En el caso que el investigador quisiera alimentar al sistema con su feedback, este se almacenará y se tendrá en cuenta en futuras recomendaciones. La siguiente ventana muestra un feedback que se ha realizado de forma exitosa:

Gracias por su participación



La información se envió correctamente

Su feedback es importante para alimentar el sistema de recomendación.



En caso de que el investigador trate de realizar más de un feedback el sistema lo detectará y le mostrará el siguiente mensaje:

Ha ocurrido un error



Su feedback no ha podido registrarse o ya se registró anteriormente.



Anteriormente, se ha mencionado un subproceso que hemos llamado "recordatorio de configuración de perfil", este subproceso es el encargado de notificar a aquellos investigadores que no han configurado su perfil y tampoco el sistema ha podido configurar un perfil base, la siguiente imagen muestra un ejemplo de correo recibido:

Hola Jose Manuel Bernabe Murcia,

Le recordamos que puede configurar su perfil para recibir convocatorias personalizadas que esten disponibles en el SGI.

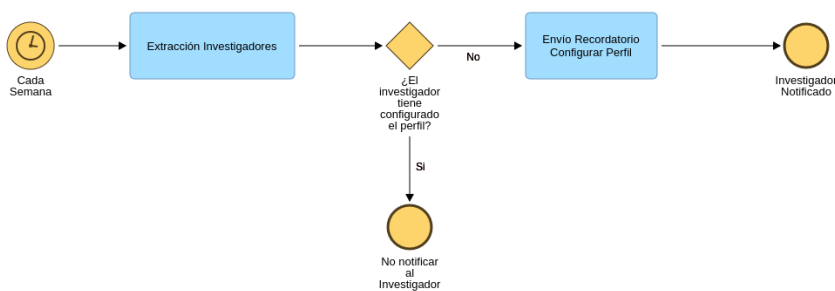
Configurar Perfil

[Configurar mi perfil](#)

****No comparta ningún enlace, son exclusivos para cada investigador.****

® Hercules-RPA

Podemos definir su BPMN de la siguiente forma:



La ventana de configuración de perfil mostrará al investigador el conjunto de áreas temáticas que hay dentro del SGI, de esta forma se puede establecer un perfil al investigador del que partir, sin la necesidad de enviar convocatorias que son totalmente opuestas al investigador. La siguiente imagen muestra la ventana de configuración de perfil:

Profile Recommendation

Áreas Temáticas

- ↳ Estructura de áreas y paneles científico técnicos de la Agencia Estatal de Investigación
- ↳ Convocatoria PCI
 - ↳ Ciencias, Ingeniería, Matemáticas, Física, Química
 - Ciencia y Tecnologías Químicas: ★ ★ ★ ★ ☆
 - Energía y Transporte: ★ ★ ★ ★ ☆
 - Ciencias sociales y humanidades: ★ ★ ★ ★ ☆

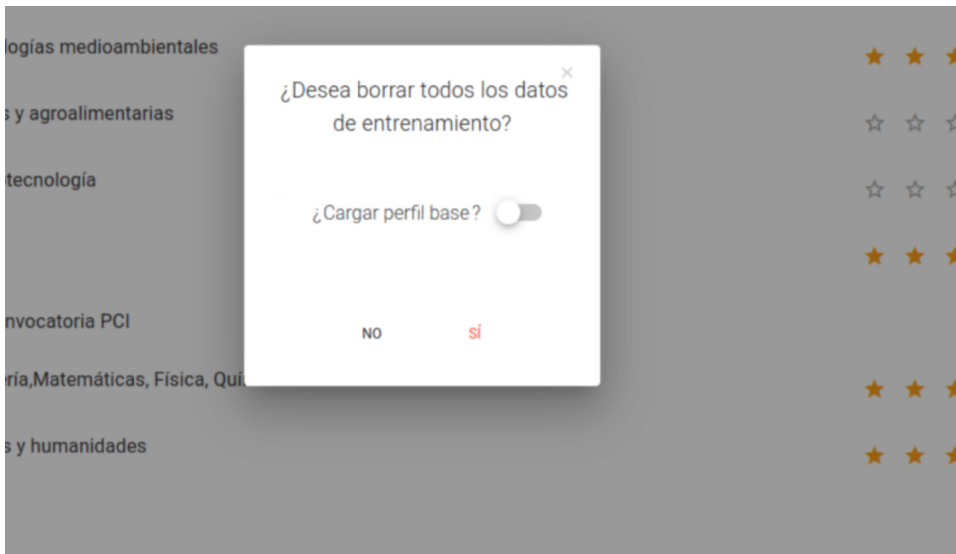
Rating : [] Tree : []

Tree : []
Tree : [object Map Iterator]
Tree : [object Array Iterator]

LIMPIAR BORRAR DATOS DE ENTRENAMIENTO ENVIAR ➤

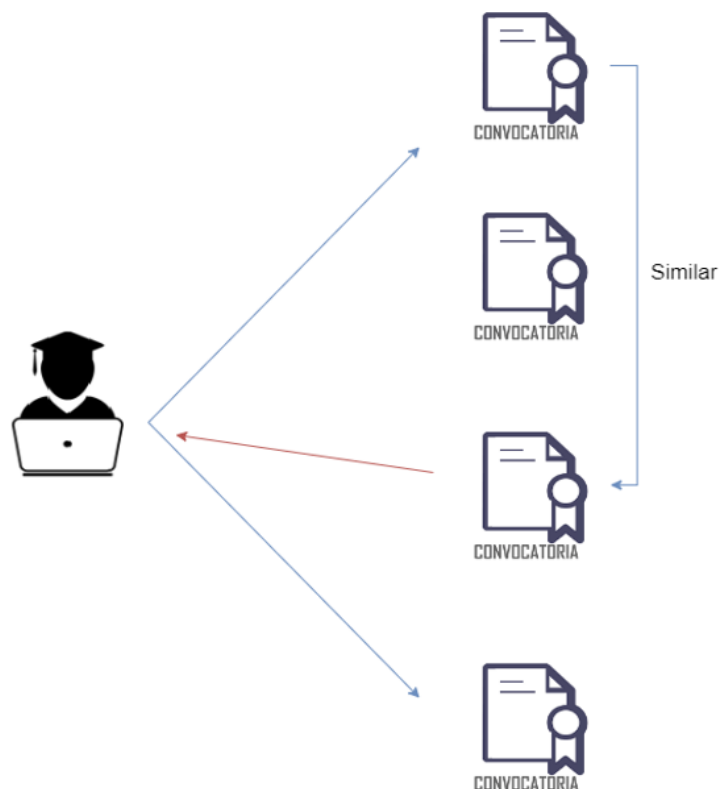
Como se puede observar existen 3 botones (la parte que aparece de Tree, son debugs, no se deben tener en cuenta):

- **Limpiar:** Vuelve a poner las puntuaciones como se habían mostrado al principio de abrir la ventana.
- **Borrar datos de entrenamiento:** Borra todos los datos del investigador que el sistema utilizaba dando la opción de volver a cargar el perfil base. Podemos verlo en la siguiente imagen:



- **Enviar.** Enviamos los datos al sistema, para que sean almacenados.

Sistema de recomendación.



El sistema de recomendación intentará personalizar al máximo lo que ofrece a cada investigador. Esto solo será posible cuando se hayan generado suficientes datos como para poder utilizar el motor de recomendación y obtener una buena tasa de acierto.

Existen diferentes tipos de motores de entre los que cabe destacar:

- **Basado en contenido:** a partir de convocatorias solicitadas, intentar ofrecer convocatorias similares.

- **Filtro colaborativo:** más novedoso que el anterior, utiliza la información de todos los investigadores para identificar similitudes. Dos grandes tipos: basado en usuarios (investigadores) y basado en ítems (convocatorias).
- **Híbridos:** los sistemas de recomendación híbridos tratan de solventar las desventajas que tienen los diferentes motores tratando de “fusionarlos”.

Para la realización de este proceso se va a **crear un motor de recomendación híbrido, fusionando el sistema basado en contenido y filtro colaborativo** de esta forma el proceso de recomendación es más completo y personalizado, ya que la desventajas de un sistema serán cubiertas por las ventajas del otro.

Filtro colaborativo: basado en ítem

Un filtro colaborativo trata de buscar similitud, cuando es basado en ítems, esa similitud se traduce a buscar ítems similares, en este caso, convocatorias. La decisión de optar por un filtro colaborativo basado en ítem y no en usuarios (similitud entre usuarios) es porque el basado en ítem se comporta mejor en muchos aspectos, por ejemplo, los investigadores pueden cambiar con el tiempo sus preferencias, pero las convocatorias serán de un determinado tipo.

La similitud entre convocatorias se buscará utilizando las áreas temáticas. Para ello, se tendrá en cuenta la puntuación/feedback que el investigador nos devuelve sobre una convocatoria, la cual se compone de diferentes áreas temáticas y que el investigador ha puntuado entre 1 y 5.

Por tanto, se puede descomponer de la siguiente forma la construcción de este motor de recomendación:

1. Identificar si el área o áreas están calificadas por, al menos, la mitad de los investigadores. En caso de no estarlo, por heurística, se podrá determinar a quién enviar esa convocatoria, teniendo en cuenta que la puntuación sea mayor de 2.5 (siempre en base a una puntuación de 1 a 5) o si ese investigador no ha calificado nunca una de las áreas temáticas.
2. En caso de que se disponga de las puntuaciones necesarias, se creará una matriz de correlación entre todas las áreas temáticas.
3. Se obtendrán los investigadores y, por cada área que hayan puntuado, se seleccionarán todas las áreas similares (según nuestra matriz de correlación), se multiplicará esa puntuación por el valor de correlación, de esta forma se potencia el valor de correlación, tanto positivamente como negativamente.
4. Se observa el resultado que se ha obtenido del área o áreas de entrada de la convocatoria, si se encuentra por encima de la media de recomendación, significa que es candidata de recomendar esa convocatoria.

¿Por qué puntuaciones de áreas temáticas y no en la propia convocatoria?

Las convocatorias tienen un periodo de solicitud, lo que da lugar a que puedan no existir suficientes solicitudes ni convocatorias para satisfacer el sistema de recomendación y que proporcione una buena probabilidad de acierto. Puede ocurrir que se tenga suficiente puntuación una vez que se haya terminado el plazo, por otro lado surge el problema de una convocatoria nueva no se sabría a quién mandarla. Se debería enviar de forma indiscriminada siempre. Las áreas temáticas, perdurarán en el tiempo y aunque aparezcan nuevas, los investigadores podrán ir puntuando esas áreas conforme vayan saliendo nuevas convocatorias, con lo que llegará un punto en que se satisfaga el requisito de mínima puntuación para utilizar el sistema de recomendación, por lo que, cuando aparezcan nuevas convocatorias, se puede determinar a quién mandarla de forma personalizada. Siguiendo este enfoque, se encontrarán relaciones entre distintas áreas temáticas que podrán ser de interés para los investigadores.

Sistema basado en contenido

Los sistemas basados en contenido tienen en cuenta características de la convocatoria, por ejemplo, áreas temáticas, unidad de gestión, modelo de ejecución, palabras clave. Todos aquellos atributos que puedan ser importantes para un investigador a la hora de tomar la decisión de optar por una convocatoria o por otra.

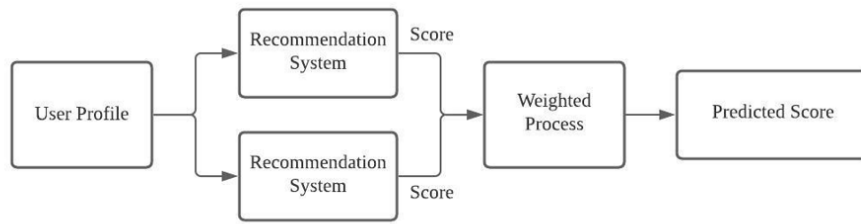
Este sistema lo que haría sería crear una matriz de similitud de las convocatorias teniendo en cuenta sus atributos, que antes de calcular la similitud se juntarían en una única columna para posteriormente vectorizarlos, y con datos suficientes de los investigadores, determinar si la convocatoria a ofrecer es similar a convocatorias anteriores. De esta forma, es posible personalizar aún más las convocatorias a notificar.

Este sistema tiene como inconveniente la dificultad para determinar qué atributos son relevantes para tener en cuenta, por lo que hace falta un experto que lo determine. Desde la UMU se han determinado que los campos a extraer para realizar este sistema de recomendación son:

- Áreas temáticas.
- Entidad convocante.
- Entidad financiadora.
- Finalidad.
- Clasificación de producción científica.
- Palabras clave.

Motor híbrido

Ambos sistemas si se construyen se pueden fusionar para solventar las desventajas que tienen cada uno de ellos por separado. Para ello, se tratará la puntuación que nos devuelve cada sistema para obtener un peso, y en función de ese peso, determinar si es candidata o no dicha convocatoria de entrada.



Enlaces a GitHub:

- Subprocesos y clases relacionadas con el proceso 4: <https://github.com/hercules-rpa/orquestador/tree/main/model/process/Process4>
- Proceso para enviar el correo de configuración del perfil del investigador: <https://github.com/hercules-rpa/orquestador/blob/main/model/process/ProcessProfileReminder.py>
- Proceso que realiza el encaje entre convocatorias e investigadores: <https://github.com/hercules-rpa/orquestador/blob/main/model/process/ProcessRecommendationSystem.py>