Projecte TERMMED. Evolución del conocimiento científico en medicina: El cambio léxico y semántico (FFFI2017-88100-P)

Acta de la reunión 4 (Aprobada 1/2/2021)

Fecha: 16 de diciembre de 2020
Hora: 16:00
Via telemàtica: meet.google.com/czx-yepg-vez

Asistentes: Mariona Arnau, José Castaño, Ona Domènech, Rosa Estopà, Judit Feliu, Fabiana da Silva França, Alejandra López Fuentes, Mercè Lorente, Zebing Lu, Jorge Porras, Òscar Pozuelo, Isabel Santamaria, Ana Rosa Terroba, Andrés Torres, Laia Vidal, Jordi Vivaldi, y dos colaboradores del HIBA. Juan Manuel Torres-Moreno ha intentado conectarse pero ha tenido problemas para seguir la reunión.
Se ha incorporado al final de la reunión: Yingfeng Xu.
Se ha excusado: M. Teresa Cabré, Amor Montané, Mariona Reverté.
Damos por excusados, por posibles incompatibilidades horarias: Daniel Luna, Coralie Schneider.

Orden del día:

1. Aprobación del acta de la reunión anterior (15/7/2020).
2. Informaciones de gestión y de las tareas realizadas.
3. Respuesta sobre la participación en los paquetes de trabajo/análisis de los datos.
4. Propuesta de calendario y formato de resultados.
5. Debate abierto.

Desarrollo de la reunión y acuerdos

Sin modificaciones, se aprueba por asentimiento el acta de la reunión anterior, enviada con la convocatoria.

M. Lorente propone un cambio en el orden del día, de manera que A. Torres pueda exponer el trabajo que ha hecho de visualización de los datos resultantes de las extracciones, ya que debe marcharse antes por estar dentro de su horario laboral. A. Torres expone la aplicación realizada y su funcionamiento, y los asistentes le dirigen algunas preguntas. Los datos que se visualizan corresponden a las 100 unidades más frecuentes y con mayor ponderación de cada período extraídas por Terminus 2.0. Se acuerda que M. Lorente pondrá a disposición de todo el equipo la aplicación HTML en el Drive compartido, para que los investigadores puedan utilizar la visualización para ilustrar sus análisis.

R. Estopà consulta si esta aplicación nos serviría para visualizar la aparición y la desaparición de formas a lo largo de los diferentes períodos. Ante la respuesta negativa,

se abre un debate sobre la manera más fácil de hacer este contraste. J. Porras propone hacerlo a partir del mismo archivo de Excel. Se acuerda que J. Porras hará este contraste y lo compartirá con todo el equipo. A partir de estos nuevos resultados, se acuerda también que A. Torres analizará si se pueden proponer otro tipo de visualizaciones contrastivas, como por ejemplo un término o más de un término en una línea del tiempo.

O. Domènech solicita que se aclare qué significa en los datos del Excel el valor de ponderación. J. Vivaldi comenta que está bien explicado en el artículo de Nazar & Cabré sobre la extracción de Terminus 2.0. Se acuerda mandar a los investigadores este artículo y el fragmento sobre la ponderación para poder entenderlo y poder citarlo, si acaso, en los análisis.

M. Lorente informa brevemente de las vicisitudes que han ocurrido estos meses que han provocado un nuevo retraso en el calendario previsto, pero que no han impedido realizar casi todas las tareas que nos habíamos propuesto en la reunión anterior (ver el punto final del acta aprobada). R. Estopà comenta que también ha sido un año complicado para los expertos con el covid, por la situación de pandemia. La única tarea pendiente es la aplicación del detector de neología en el corpus del proyecto. Se acuerda mandar un recordatorio a A. Torres y a J.M. Torres-Moreno.

R. Estopà comenta que sería conveniente compartir una misma descripción del corpus para todos los paquetes de trabajo. Se propone partir del texto que se enviará, antes de Navidad, como propuesta al XII Congreso Internacional de Lingüística de Corpus (CILC2021), que se celebrará telemáticamente en abril de 2021, organizado por la Universidad de Murcia y la asociación AELINCO. M. Lorente y R. Estopà pondrán a disposición de todos los investigadores esta descripción antes de la próxima reunión.

Alejandra López expone lo que se ha realizado con la información del corpus: la tabla de Excel, los metadatos de cada texto, el conteo de los token de cada texto, y el traspaso de la información a un archivo de textos. Indica que se han realizado los conteos mediante dos programas distintos (uno ella misma y el otro Oscar Pozuelo) y se ha visto que los resultados no son del todo coincidentes, debido a los algoritmos de cada programa. En algunos casos la diferencia es mínima, en otros casos es considerable, seguramente porque uno de los programa no incluye el texto de las tablas. Se acuerda que en el Excel y en el archivo de textos resultante se recojan los dos conteos con indicación en los metadatos de cada texto de qué programa se ha aplicado. A. López explica el estado de la cuestión del archivo de textos (pendientes aun de traspasar los datos de 2 períodos), con el compromiso de terminar el trabajo esta misma semana. Se acuerda poner a disposición de todos los investigadores la lista final de los textos del corpus.

Jorge Vivaldi expone lo que ha hecho en extracción automática. Reproducimos sus resúmenes enviados por correo electrónico, que coinciden con su exposición:

> *Ya tenemos a punto un fichero que es la combinación de Terminus, Yate y WikiYate. He creado el directorio TermMed->extraccio_tgica_Combinacio donde hay el resultado de la combinación de los tres extractores para cada periodo. Es un fichero de texto con tres columnas:*
>
> 1. *Detección: qué extractores han detectado cada candidato a término (T: Terminus, Y: Yate i W: WikiYate)*
> 2. *Coeficiente de terminologicidad: es un valor entre 0 y 12. Cada extractor puede asignarle un valor entre 0 y 4. En esta columna hay la suma de los tres resultados. En cada extractor hay diferentes métodos de validación (excepto Terminus) y cada uno aporta como máximo 1 punto.*
> 3. *CAT: candidato a término.*
>
> *Si fuera necesario puedo añadir más información (frecuencia, resultados individuales, ficheros en los que aparecen, etc). La coincidencia entre YateMCR y Terminus es de 834 candidatos En el caso de WikiYate tengo un problema con la codificación de los caracteres que espero resolver pronto.*

M. Lorente aprovecha la ocasión para comentar la falta de coincidencia entre los resultados de extracción de Terminus y de Yate, que J. Vivaldi, después de haberlo hablado con A. Montané, indica que hay que buscarla en las diferentes estrategias de detección que tienen ambas herramientas. Se acuerda que los resultados de extracción y los comentarios sobre las distintas estrategias sean objeto también de un informe, y posibles publicaciones derivadas.

Se abre la revisión del documento de asignación de paquetes de análisis que se mandó con la convocatoria de esta reunión. Después de las contribuciones de los asistentes y del debate sobre las posibilidades de cada paquete de trabajo, R. Estopà recuerda que el TFM de Mariona Reverté también incluye estudio de webs. Se acuerda que R. Estopà y M. Lorente actualizarán el documento y lo volverán a mandar a todos los investigadores, con el acta de esta reunión.

R.Estopà interviene para exponer la previsión de resultados del proyecto, atendiendo al hecho de que se cierra inicialmente (informe final al ministerio) a finales de marzo de 2021 y que podremos acabar de publicar resultados a lo largo del 2021.

1) Informes tangibles de cada paquete de trabajo para el informe final y para publicarlos en abierto en un e-repositorio con DOI.
2) Propuestas de comunicaciones en congresos internacionales y de artículos en revistas, para ver si recibimos a tiempo respuestas positivas.
3) Planificación de la edición de un libro con los informes de todos los paquetes de trabajo y una síntesis de resultados.

M. Lorente, en la misma línea, propone el calendario a seguir:

- **Antes de Navidad 2020**: Se enviaran las nuevas listas de apariciones y desapariciones del Excel de resultados, la lista de textos contenidos en el corpus con descripción, la nueva propuesta de paquetes de trabajo con indicación del coordinador de cada paquete.
- **20 de enero de 2021**: Reunión del proyecto con un preanálisis de los datos de cada paquete. Debate y resolución de dudas.
- **Mediados de febrero de 2021:** Reunión con los análisis terminados. Exposición de resultados.
- **Mediados de marzo de 2021:** Reunión con los informes redactados. Preparación de la memoria final.
- **Abril de 2021**: Reunión de cierre. Previsión de publicaciones. Acciones de difusión de resultados.

Se acuerda por asentimiento este calendario. Se fija la próxima reunión para el miércoles 20 de enero a las 16h.

En el turno abierto, A.R. Terroba plantea que se podría aprovechar el plan de La Rioja para divulgación de resultados de proyectos científicos para difundir nuestro proyecto y sus resultados.

Los participantes de la reunión comentan que, a pesar de la situación de pandemia, hay que alegrarse de los avances actuales y futuros del proyecto, y intercambian los mejores deseos para el cambio de año .

Se cierra la reunión a las 17h45.