

Integración en la plataforma RESCATA de analizadores AMR existentes

Un paso fundamental para conseguir construir la plataforma RESCATA es crear un analizador AMR (Abstract Meaning Representation en inglés o Representación Abstracta del Significado) para textos completos. La propuesta de RESCATA aprovecha las herramientas AMR existentes que trabajan a nivel de oración (sentence-AMR) para conseguir la representación AMR del texto completo (text-AMR). Por ello, se analizaron 4 analizadores AMR creados en la tarea de evaluación 8 del SemEval en 2016 y 2017. A continuación, se describe brevemente cada uno de ellos, así como posibles problemas:

*msrsplat*¹ es un analizador lingüístico y estadístico de microsoft que incluye análisis lingüístico tradicional, como análisis morfológicos, y desarrollos más recientes. Entre estos últimos se encuentra el analizador de AMR para 5 lenguas (inglés, español, alemán, francés y japonés). El analizador AMR está disponible para su uso como servicio web REST, pero en ocasiones está caído por lo que disponer del grafo AMR no siempre es disponible.

*amr-eager*² es un analizador para lengua inglesa se basa en transiciones para analizar AMR. Se encuentra disponible como una herramienta independiente lista para instalar en linux. Sin embargo, sus requisitos no pueden ser instalados correctamente en la distribución de nuestros servidores linux. Por lo tanto, no es posible obtener un grafo AMR.

*camr*³ es un analizador de AMR basado en transiciones de árboles y grafos para lengua inglesa. A pesar de que se encuentra disponible como una herramienta independiente y de fácil instalación, no es un sistema de producción pues cuando se produce algún error abre el depurador y no produce el grafo AMR.

*jamr*⁴ es un analizador semántico de AMR comúnmente utilizado, que además incluye módulos para generación y alineación. Este último módulo es empleado en otras herramientas como *camr*. Se encuentra disponible como una herramienta independiente de sencilla instalación en linux que sí nos ha permitido producir grafos AMR.

La siguiente tabla ofrece los resultados que cada una de ellas obtuvo durante la competición SemEval, así como un resumen de las dificultades detectadas. Finalmente, dados los resultados que consiguieron estos analizadores y la disponibilidad para obtener el análisis AMR, nos decidimos por integrar *jamr* en nuestra plataforma como analizador AMR a nivel de oración.

¹ <http://research.microsoft.com/msrsplat>

² <https://github.com/mdtux89/amr-eager>

³ <https://github.com/c-amr/camr>

⁴ <https://github.com/jflanigan/jamr>



rescata
Representación Canónica
y Transformaciones
de los textos aplicado a las
Tecnologías del Lenguaje Humano

Herramienta	Resultados en LDC2015E86 test set	Comentario
mrsrplat	F 50.4	Servicio en ocasiones caído.
amr-eager	F 64	Requisitos no instalados correctamente en nuestra distribución de linux.
camr	P 71 R 63.89 F1 67.33	No es una herramienta para sistemas en producción, las excepciones abren el depurador.
jamr	P 70 R 65 F 67	Herramienta que ofrece más disponibilidad y mejores resultados.