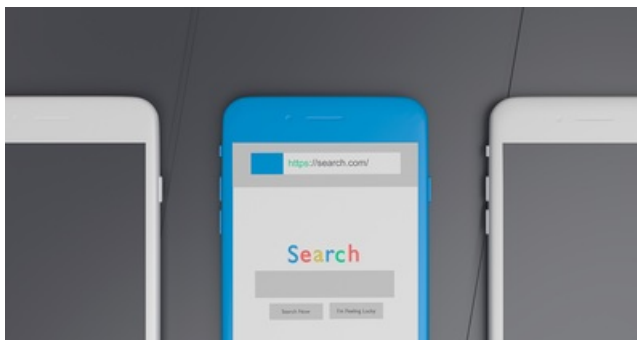


Novos métodos de fusión de rankings melloran o rendemento dos sistemas de procura

luns, 19 febreiro, 2018 - 13:02

O traballo, publicado na revista 'Information Fusion', propón melloras na construción dos 'benchmarks' utilizados polos buscadores en liña.



Os *benchmarks*, ou «banco de probas», son ferramentas para medir e comparar a efectividade dos algoritmos e sistemas de ranking de documentos que empregan as tecnoloxías de busca dispoñibles na actualidade (procura web, buscadores verticais, etc.). Estes conxuntos de probas son un activo fundamental no desenvolvemento das tecnoloxías de busca, e ao longo de décadas tiveron unha importante incidencia, non só a nivel comercial (con exemplos próximos nos motores de procura web máis populares, como *Google*, *Bingo* u *Yahoo*) senón tamén a nivel académico, con numerosos avances científicos no eido da *Recuperación de Información*.

Estes benchmarks ou conxuntos de 'test' están formados por bases documentais inxentes, entre as que se inclúen millóns de documentos, consultas de usuario, e información de relevancia (é dicir, os documentos de meirande importancia en cada tipo de consulta). Polo xeral a 'etiquetaxe de relevancia' require da intervención de expertos humanos, o que permite establecer un "estándar de ouro" de grande utilidade para a comunidade científica. Con todo, o tamaño destas coleccións fai inviable a súa análise completa por parte dos expertos, sen esquecer que a maioría dos textos non adoitan ser relevantes para consultas específicas. Todas estas dificultades teñen convertido en práctica común que a 'etiquetaxe de relevancia' se materialice a partir de tan só unha mostra de documentos para cada consulta do *benchmark*.

Agora, un novo traballo do CiTIUS publicado pola revista científica *Information Fusion* presenta unha nova técnica que mellora substancialmente estes procesos de etiquetación de relevancia. Dirixido polo investigador adscrito ao centro **David E. Losada**, o estudo explora a aplicación de distintos **métodos de fusión de rankings** para determinar que documentos deben ser avaliados en cada consulta. O artigo desenvolve así unha comparación exhaustiva de diferentes técnicas, demostrando que a metodoloxía proposta pode mellorar a calidade dos *benchmarks* construídos e, ao mesmo tempo, reducir notablemente o esforzo necesario para a súa construción.