

لینک گیت‌هاب:

<https://github.com/Armin11DSE/Web-Information-Retrieval-Projects>

برای افزایش سرعت جست‌وجو از LSA (Latent Semantic Analysis) استفاده شده است، که نتیجه جست‌وجو را با استفاده از context بهبود می‌بخشد. کاربردهای LSA:

<https://aioseo.com/seo-glossary/latent-semantic-analysis/#:~:text=Search%20engines%3A%20LSA%20helps%20improve,content%20they%20are%20currently%20viewing>

توضیحات کلی درباره کد:

نتایج جست‌وجو با استفاده از cosine-similarity و TF-IDF (Term Frequency – Inverse Document Frequency) به دست آمده‌اند؛ سپس با استفاده از دریافت بازخورد و استفاده از الگوریتم Rocchio نتایج اولیه بهبود یافته‌اند و مشاهده می‌شود که نتایج به هدف ما نزدیکتر خواهند بود. نتایج کلی:

بدون بازخورد، سیستم صرفاً بر اساس شباهت ظاهری بین کلمات عمل می‌کند و نتایج ممکن است حاوی نویز باشد. با اعمال الگوریتم Rocchio و استفاده از بازخورد کاربر، سیستم می‌تواند ویژگی‌های مشترک بین اسناد مرتبط را استخراج کند و نتایج را بر آن اساس بهبود ببخشد، این باعث می‌شود اسناد مرتبط به رتبه‌های بالاتر منتقل شوند و Precision افزایش یابد.

* تمامی نتایج برای کوئری "اقتصاد" بوده‌اند و اسناد مرتبط ۱، ۲، ۴ در نتایج بوده‌اند.