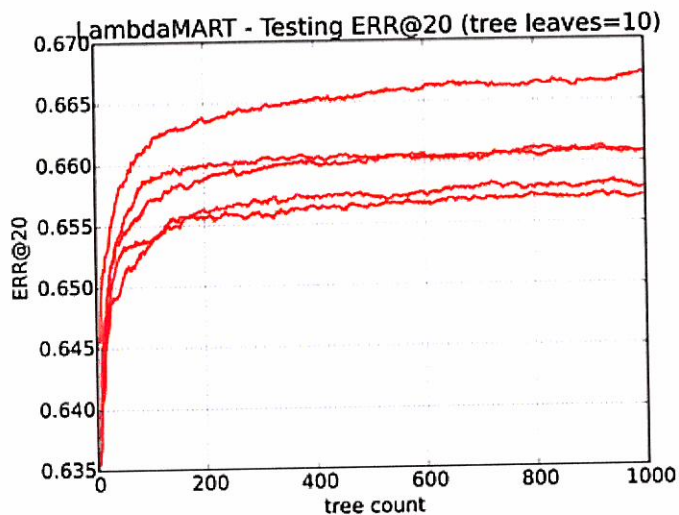


Learning to Rank

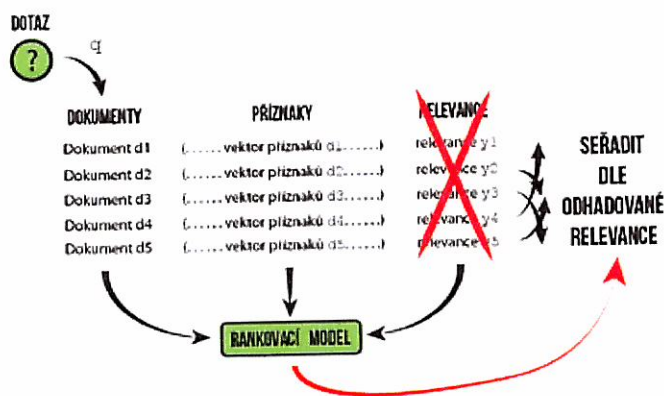
Řešitel: Bc. Marek Modrý, FEL ČVUT

Vedoucí projektu: Ing. Jan Šedivý, CSc., FEL ČVUT

Fenoménem současné doby je rapidní nárůst množství sbíraných dat. Odborníci navíc odhadují, že v následujících letech se toto množství může ještě více než stonásobně zvýšit.



Záplava nových dat však vytváří tlak na to, abychom byli schopni v dostupných informacích hledat to, co je pro nás relevantní. *Learning to Rank* je disciplína strojového učení, která při využití dostupných dat a dotazů vytváří matematický model využitelný pro následné řazení informací (dokumentů) podle jejich relevance. Využití algoritmů *Learning to Rank* je široké, a to od internetových vyhledávačů přes doporučovací systémy, opravy překlepů, online reklamy až například po přiřazování autorů k článkům v Microsoft Academic Search a mnoho dalších možností.



Cílem projektu bylo pečlivě projít a analyzovat současné „state-of-the-art“ algoritmy a vedle nich pak prozkoumat používané metriky pro měření kvality naučeného modelu a dostupné datové sady, které by algoritmus mohl použít k učení. Během projektu bylo analyzováno 17 algoritmů, a to včetně jednoho, který do projektu laskavě poskytla společnost Seznam.cz. Dále pak bylo získáno a analyzováno 15 dostupných datových sad, přičemž největší obsahovala 1 200 192 dokumentů. V neposlední řadě bylo porovnáno osm hodnotících metrik. Výstupem projektu byla diplomová práce obsahující výsledky nejdůležitějších experimentů a vysvětlení pozorování u jednotlivých datových sad a metrik kvality řazení. Diplomová práce byla následně využita společností Seznam.cz jako podklad pro analýzu kvality jejího stávajícího řešení v porovnání s aktuálními veřejně známými „state-of-the-art“ algoritmy. Dále pak byl v rámci projektu implementován nejlépe hodnocený algoritmus. Implementováno bylo navíc i několik modulů umožňujících vizualizaci průběhu učení modelu v čase. Projekt také poukázal na několik možných vylepšení a potenciálních nových problémů, jejichž vyřešením by bylo pravděpodobně možné zlepšit kvality některých algoritmů a dostupných datových sad. Výstup projektu vyústil v další spolupráci se společností Seznam.cz.

