**Zhodnocení projektu „Vývoj technologie rozpoznávání a indexování obsahů monografií“**

V rámci projektu byla vyvinuta metoda na vytěžování klíčových slov z knižních obsahů pro účely indexování v discovery systémech či jiných vyhledávačích, aby bylo dosahováno vyšší relevance při vyhledávání podle věcných hledisek.

Výstupem projektu je webová služba spolu s pracovním postupem, které automatickou tvorbu klíčových slov z obsahů umožňují. Obsahy knih jsou nejdříve naskenovány a zpracovány programem pro optické rozpoznávání znaků, jehož výstupem jsou jednotlivá slova tak, jak se vyskytují v obsahu. Tato slova jsou následně morfologicky analyzována a každému z nich je přiřazen slovní druh a základní slovní tvar (lemma). Autosémantická slova jsou odstraněna a všem ostatním je přiřazeno skóre, které odpovídá jejich významnosti. Výstupem procedury je množina slov s nejvyšší významností (tedy klíčová slova). Metoda má několik parametrů, kterými lze ovlivnit množství slov na výstupu ( min. počet, max. počet, min. hodnotu významnosti). Významnost slov je odhadována jednak na základě počtu výskytů v obsahu zpracovávané knihy, jednak na základě počtu výskytů v referenčním korpusu.

Webová služba nyní běží v pilotním režimu a probíhá na ní v rámci testování zpracování první ostré dávky knižních obsahů v objemu 100 kusů. Na konci měsíce ledna bude chování služby vyhodnoceno a dojde k její optimalizaci dle zjištěných nedostatků. Následně bude služba uvedena do provozního režimu jako webová služba, kterou bude možné externě využívat.

Webová služba je vytvořena v jazyce python.

Na řešení projektu se Národní technická knihovna podílela společně s Univerzitou Karlovou v Praze.

Přidělené prostředky byly využity tak, jak bylo v podmínkách dotace uvedeno.