



Mapování znalostí prostřednictvím indexování publikační aktivity

Josef Nožička, Petr Jirků VŠE Praha  **Aquasoft**

Obsah

- Cíl mapování znalostí
- Východiska
- Naš přístup
- Znalostní mapa
- Postup mapování znalostí
- Statistiky
- Ověření spolehlivosti statistik
- Výsledky
- Nové statistiky
- Objektivizace vyhodnocení modelu
- Odlišení osob s různou mírou znalosti
- Závěr

Cíl mapování znalostí

- Vychází z praktické potřeby větší SW firmy
- Umožnit uživatelům systému zodpovídat dotazy typu:
 - „Kdo je expertem na standard UML?“
 - „Je u nás někdo, kdo už se setkal s technologií ASP.NET?“
 - „Kdo vytvářel v roce 1999 bezpečnostní systém pro intranetové aplikace?“
- To je umožnit organizaci vytvořit si přehled o znalosti, kterou má... a následně ji využít
 - Kolegy (spolupráce na řešení problému, úvod do problematiky)
 - Personalisty/Řízení (sestavování týmů, identifikace znalostních deficitů)



Východiska

- Organizace postavená na expertech
 - Nová řešení vznikají interakcí jednotlivých lidí zodpovědných za vyřešení problému
 - Nová řešení nelze plánovat, ale lze je stimulovat vhodným prostředím.
- Minimalizace zátěže expertů – oproti obdobným přístupům (expert location)
 - Není třeba udržovat taxonomii jednotlivých oblastí expertízy
 - Není třeba udržovat znalostní profily
 - Systém otevřený i pro znalosti nepodchycené v okamžiku tvorby znalostní mapy ... Paralela k tacitním znalostem.
- Respektování soukromí zaměstnanců

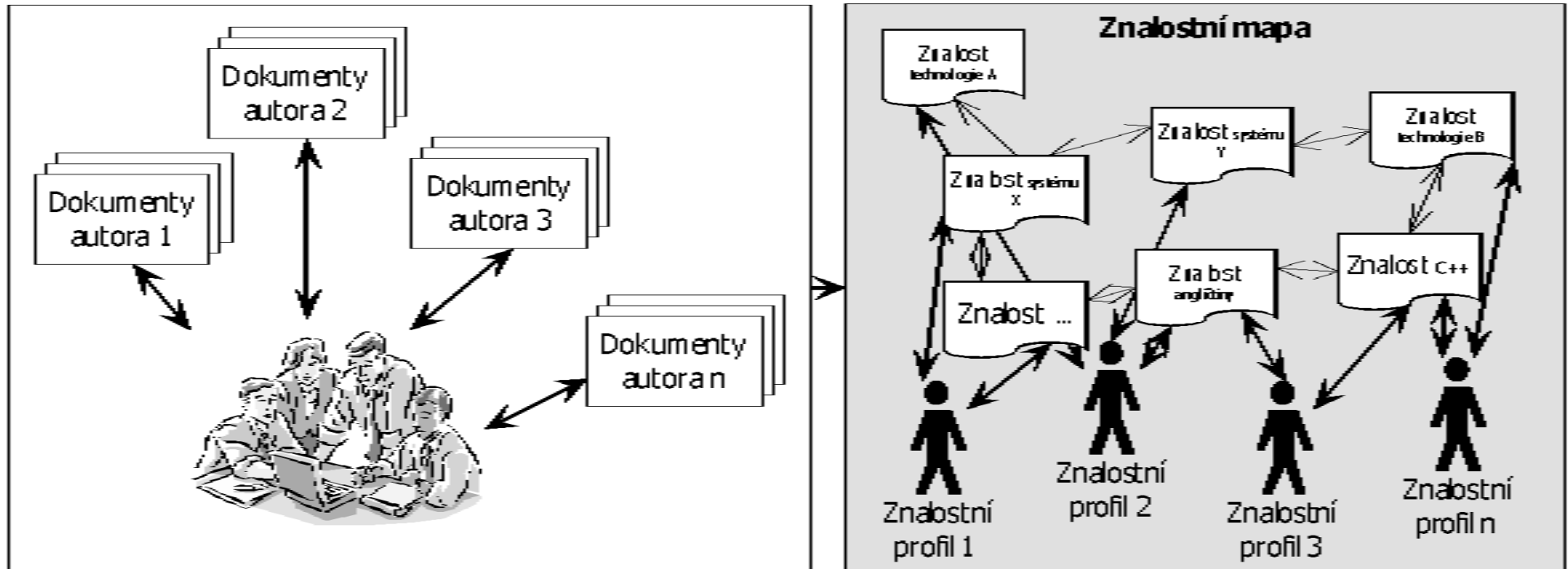


Náš přístup

- Vycházíme z dokumentů uložených ve firmě
 - Jde o analytické dokumenty obsahující rozборы věcné problematiky, právní dokumentaci, zadání pro software
 - Obsahující informaci o autorovi a datu uložení
 - Dokumenty nejsou nijak pevně strukturované
 - Vytipování adresářových struktur obsahujících důležité dokumenty
 - Dokumentové a verzovací repository
- Podstata přístupu
 - Podchytit přirozenou dynamiku lidské znalosti a využít ji k odhadu aktuálního stavu znalostí dané osoby
 - Využití informace o časové složce k odhadu hloubky znalosti dané osoby



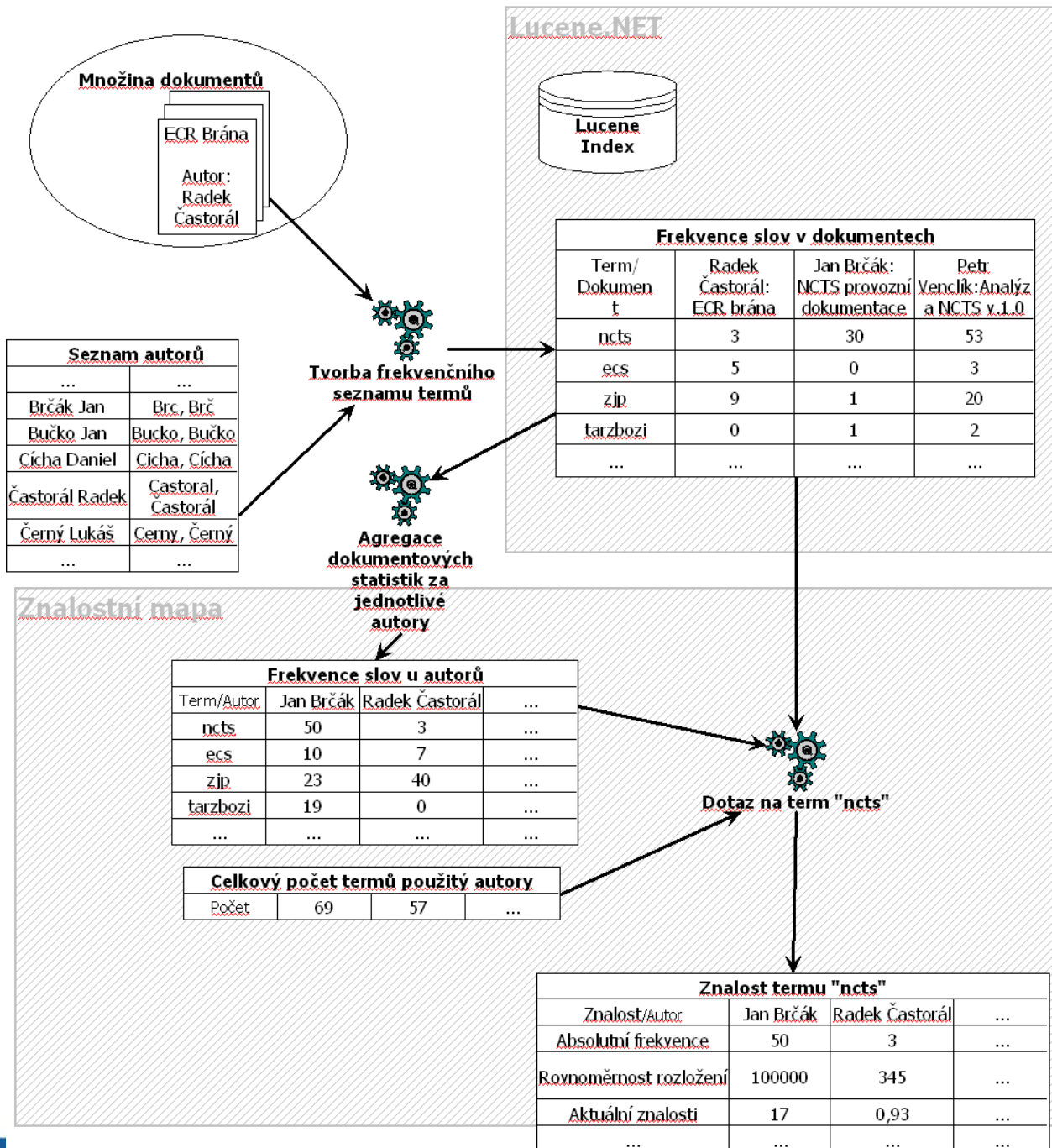
Znalostní mapa



Znalostní mapa

- Znalostní profil
 - Sada dokumentů vytvořená daným autorem
 - Sada dokumentů s jiným vztahem k autorovi
- Ve fulltextovém indexu je každý dokument
 - Opatřen informací o časové složce
 - Informací o autorství
- Sada algoritmů pro tvorbu statistik
 - Modifikace dle míry publikační aktivity autorů
 - Scoring dle počtu výskytů, scoring dle vyhledávacího nástroje
- Technologie
 - MS .NET Framework
 - Lucene.NET

Postup mapování znalostí



Statistiky

- Statistika ignorující časovou složku
 - Počet výskytů (1)
- Statistiky pracující s rozložením užití termů v čase
 - Rovnoměrnost rozložení (2)
 - Rovnoměrnost rozložení vážená (3)
 - Znalostní následnost (6)
- Statistiky upřednostňující aktuální znalosti
 - Aktuální znalosti (4)
 - Aktuální znalosti vážené (5)



Ověření spolehlivosti statistik

- Testovaná data
 - Čítá v současnosti 12955 dokumentů od 121 autorů
 - Kompletní analytická dokumentace jedné z divizí firmy
- Metoda ověření
 - 21 vybraných termů (neskloňovatelné, věcné i technické povahy užívaných divizí) u nichž je možné jednoznačně určit hlavního nositele
 - Subjektivní ohodnocení míry znalosti
- Výsledek směl nabývat 3 hodnot
 - Daná statistika správně identifikovala alespoň v první pětici nositelů znalostí a správně určila i pořadí
 - Daná statistika správně identifikovala alespoň v první pětici nositelů znalostí
 - Daná statistika neidentifikovala korektně hlavní nositele znalostí



Výsledky pro dokumenty v proj. adresářích

■ Scoring dle Lucene .NET

Výsledek/statistika	Počet vyskytů (1)	Rovnoměrnost rozložení (2)	Rovnoměrnost rozložení vážená (3)	Aktuální znalosti (4)	Aktuální Znalosti vážené (5)
Pořadí autorů	38,10%	47,62%	47,62%	38,10%	42,86%
Identifikace autorů	85,71%	85,71%	85,71%	80,95%	85,71%

■ Prosté sčítání výskytů

Výsledek/statistika	Počet vyskytů (1)	Rovnoměrnost rozložení (2)	Rovnoměrnost rozložení vážená (3)	Aktuální znalosti (4)	Aktuální Znalosti vážené (5)
Pořadí autorů	57,14%	61,90%	71,43%	38,10%	52,38%
Identifikace autorů	95,24%	100,00%	100,00%	95,24%	95,24%



Výsledky projektové adresáře v kombinaci s verzovacím systémem (MS VS SourceSafe)

Výsledek/s tastika	Počet vyskytů (1)	Rovnoměr nost rozložení (2)	Rovnoměr nost rozložení vážená (3)	Aktuální znalosti (4)	Aktuální Znalosti vážené (5)	Znalostní následnost (6)
Pořadí auto	80,00%	93,33%	86,67%	86,67%	73,33%	93,33%
Identifikace	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Nové statistiky vycházející z teorie učení

- Harry Bahrick – Permastore
 - Věci jednou uložené v dlouhodobé paměti prakticky nemizí
- Kognitivní teorie dlouhodobé paměti – zapomínání je funkcí následujících faktorů
 - Iniciální učení
 - Počet opakování
 - Doby od posledního užití
- Cohen, Conway, Stanhope – studie na 450 bývalých studentech Open University
 - Znalosti klesají po dobu 2 let od učení
 - Zbývajících zhruba 30% znalostí již následně neklesá
- Statistiky
 - Zapomínání znalostí
 - Znalostní plocha



Objektivizace zjišťování měření výsledků

- Jak na to
 - Jak ověřit, že je někdo expertem?
 - Jak stanovit míru expertizy?
- Zvolené řešení
 - Statistický průzkum u lidí působících dlouhodoběji ve firmě
 - Zastoupení: Vedoucí projektů, team-leadeři, i jednotliví zaměstnanci
 - Ověřováno na množině 35 termů
 - Cílem je stanovit nikoli absolutní míru, ale relativní míru znalosti
- Omezení
 - Na množinu lidí, kteří produkují nestrukturované texty (analytici, vedoucí projektů, návrháři)
 - Na jednu divizi firmy (30 lidí)



Porovnávání míry znalosti

- Jak zjistit, jestli dva autoři, kteří se umístili ve skórování za sebou mají stejnou či odlišnou míru znalosti
- Vychází z měření znalostní plochy
- Metoda monte Carlo
- Měření statistické významnosti rozdílů znalostních ploch podle statistických rozložení
 - Počtu dokumentů v čase
 - Rozložení v čase
 - Skóre jednotlivých dokumentů
- Pro nejlepší autory



Závěr

- Výsledek
 - Nástroj pro vyhledávání expertů ve středně velkých firmách
 - Nenáročný pro začlenění do IS firmy
- Další postup
 - Dopočítání výsledků, ověření funkčnosti
 - Zahrnutí dalších dokumentových repository používaných ve firmě (kromě VSS i SubVersion, MS SharePoint)
 - Indexace dokumentů/přírůstková indexace
 - Nasazení do ostrého běhu v Aquasoftu

Díky za pozornost...

... a prosím o podněty, otázky.

Josef Nožička

j.nozicka@aquasoft.eu

<http://www.aquasoft.eu>