

Zápisnica zo stretnutia tímu č. 15 konaného dňa 9.10.2014

Roly priradené členom tímu pre prvý šprint:

- 1.Vedúci tímu – Martin Šustek
- 2.Manažér rozvrhu a plánovania – Ján Krivý
- 3.Manažér rizík – Vlado Demčák
- 4.Manažér podporných prostriedkov a vývoja – Marek Černák
- 5.Manažér kvality a dokumentácie – Martin Cibula
- 6.Zástupca vedúceho a hlavný architekt – Juraj Michalička
- 7.Manažment komunikácie a ľudských zdrojov – Jozef Melko

Implementačné úlohy pre aktuálny šprint:

- 1.dostaneme názov autora alebo aj ďalšie metadáta. Nimi sa dopýtame do WOS a SCOPUS a vrátíme výsledok. Systém si identifikuje, kto to je, vypýta si IDčka z WOS s SCOPUS a tie priradí k článku, ktorý citáciu použil. Jeden metazáznam pre tento článok bude ID a názov zdroja
- 2.dostaneme ID WOS-u a musíme potiahnuť všetky ohlasy, ktoré boli na dielo
- 3.všetky záznamy si budeme niekam ukladať. Na to musíme vytvoriť dátový model, ktorý bude rešpektovať spôsob prepojenia týchto entít
- 4.vytvoriť webovú službu pre získavanie IDs z WOS a SCOPUS + podľa získaných dát navrhnuť dátový model. (ak sa podarí, tak aplikovať službu, ktorá vyhľadá ohlasy)
- 5.Pre každý typ dokumentu (časopis, článok zo zborníka, kniha, elektronický dokument, ...) musíme mať príklady získaných údajov – do excelu uložiť metadáta a vedľa nich príklady – záložka pre každý typ dokumentu (potrebujeme získavať prieniky)

Pre uchovávanie záznamov treba navrhnuť 2 databázové schémy:

- DB1 – IDčka z WOS a SCOPUS
- DB2 – Ohlasy na IDčka z WOS a SCOPUS (treba deduplikovať ohlasy)

!!! Pri taskoch dokumentovať literatúru, ktorú som pri tasku použil a závery, ktoré mi z jeho riešenia vplynuli !!!

Poznámky k technológiám a postupom:

- naštudovať štandardy Unimark a Mark 21
- databázy vracajú malé objekty obsahujúce skupiny názvových údajov s vlastnosťami (názov autora, názov podčasti, ...). Sú očíslované od 0,00 po 9,99 – niečo je dohodnuté na nadnárodnej a niečo na korporátnej úrovni a tým pádom niečo sa dá stiahnuť iba z niektorých zdrojov
- z50 protokol kedysi určoval, aké atribúty a spôsob ich získavania je používaný na sťahovanie dát a to zaručuje rovnakú štruktúru dát, ktoré sú vymieňané
- TAG má indikátory (prvé dva sú vyhradené) nie vždy sú oba použiteľné a určujú, či je názov použiteľný. Nesú sémantickú informáciu, či je napr. Objekt použiteľný
- budeme hľadať štruktúru údajov knihy. Máme autorov a na základe selekčných prvkov hľadáme. Každý prvok zapíšeme inak (Pavol Návrat, Návrat P., atď.). Čiže sa stavajú súbory autorov. - unifikované (zapísané jednou formou) selekčné prvky s nevyhnutným poznámkovým a odkazovým aparátom. Selekčným prvkom napr. nie je iba Pavol Návrat, ale aj celá informácia. Pridáva sa nejaká Datácia, ktorá odlišuje duplicitné mená a podobne
- do záznamu ide aj identifikátor, na základe ktorého vieme literály identifikovať, keď sú použité vo vzťahu
- RDF – resource description framework
- VIAF.org prepája identifikátory knižníc jednotlivých krajín tak, že urobia slovenskéID IS_SAME_AS maďarskéID, čím získavame medzinárodný identifikátor
- a k čomu vedieť, čo publikoval Havel v DE:

1.zoberiem moje ID

2.dám ho do VIAF, čím získam medzinárodný záznam

3.z neho vezmem DE ID a tým sa dopýtam do DE knižníc

- Články, publikované viacerými inštitúciami vo viacerých publikáciách, sú zaradené do WOS a SCOPUS – medzinárodné indexy pre vyhľadávanie citácií
- Existujú citačné normy a štýly. Z použitej literatúry sú pomocou regexov vyťahované odcitované diela a pripájané k autorom ako ohlasy na ich publikácie
- ak ID vo WOS a SCOPUS má fulltext – treba pod ID zdroja vo filesystéme ukladať aj fulltexty (nie len záznamov, ale aj ohlasov)
- pozor na veci typu dizertačná práca – abstrakt je prístupný verejne, ale fulltext je na 5 rokov prístupný iba pre fakultu a až potom pre verejnosť
- OAIPMH – protokol pre interoperabilitu – aby nebolo nutné ručne zasahovať a niečo by s našou databázou spolupracovalo automatizovane