Hlavný cieľ aplikácie:

* Sumarizácia údajov o publikáciách a projektoch

Účel:

* transparentnosť financovania

Nástroje:

* Rozpočet – projekty, publikácie
* Periodické hodnotenie publikácií
* Rebríčky zamestnancov

Pohľady (s cieľom dosiahnutia transparentnosti môže každý zobraziť všetko):

* Jednotlivec: špeciálne: ja + ktokoľvek iný
* Pracovisko (katedra) – sumár zamestnancov aj celkový sumár
* Odbor pre periodické hodnotenie: na jednom pracovisku môže byť aj viac (informatika, učiteľstvo), jedno pracovisko môže byť zahrnuté do viacerých odborov (učiteľstvo informatiky + matematiky + …)
* Fakulta
* **Admin, ktorý nahadzuje a edituje údaje**

# Funkcionality

Projekty (moje)

Publikácie (moje)

Periodické hodnotenie (moje)

Štatistiky

* Publikácie (po rokoch) cca 4000 ročne
* Projekty (iný typ zobrazenia)
* Rozpočet

Admin – položka, ktorá pribudne ak daná rola je admin (zistí sa na základe položky z users)

* Import publikácií
* Import projektov

# Import údajov

## Publikácie

Import spočíva vo viacerých krokoch

### Import zoznamu publikácií

* txt súbor – zoznam publikácií za dvojročné obdobie – pri importe je potrebné sledovať, či daná publikácia (dané referenčné číslo) už v systéme nie je. Ak je, skontrolovať správnosť údajov, ak nesedia – dať do zoznamu úprav pre zeditovanie – tento import je hodne sprznený, je tam veľa enterov, treba ho upraviť. Oddeľovačom je text „ref. číslo:“. Keď budete parsovať jednotlivé záznamy, možno sa vám podarí využiť, že jednotlivé typy publikácií majú trocha odlišnú štruktúru položky (na začiatku som preto nechal aj zoznam typov položiek). Údaje sú v ***UKF.Nitra-2023-09-12.rar***.
* Import som robil tak, že som identifikoval publikáciu a potom ju importoval ako jeden riadok databázy a následne po importovaní všetkých som nad ňou robil experimenty)
* Obsah publikácie má podobu:

ref. číslo: 481234

**V1\_024 Media - Migration - Politics [textový dokument (print)] [elektronický dokument]**

**[editovaná kniha] : Discursive Strategies in the Current Czech and Slovak Context / Spálová,**

**Lucia [Autor, UKFFFAKMR, 10%] ; Beneš, Jan [Autor, 10%] ; Cabada, Ladislav [Autor, 10%] ;**

**Just, Petr [Autor, 10%] ; Charvát, Jakub [Autor, 10%] ; Charvát, Martin [Autor, 10%] ;**

**Charvátová, Denisa [Autor, 10%] ; Mikuláš, Peter [Autor, UKFFFAKMR, 10%] ; Niklesová, Eva**

**[Autor, 10%] ; Rosůlek, Přemysl [Autor, 10%] ; Dobek-Ostrowska, Boguslawa [Zostavovateľ,**

**editor] ; Glowacki, Michal [Zostavovateľ, editor]. – 1. vyd. – č. 17. – Berlín (Nemecko) : Peter Lang, 2022. – 245 s. [tlačená forma] [online] : text. – [angličtina]. – [OV 060]. – [ŠO 7205]. – [recenzované]. – ISBN 978-3-631-86275-9. – ISBN (online) 978-3-631-85877-6. – ISSN**

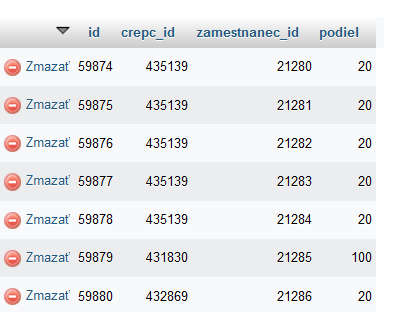
**2197-1625. – DOI 10.3726/b19495**

* Ak pri zostavovateľovi nie sú percentá, k autorom sa nepočíta, inak ho počítať ako autora
* Ak je pri autorovi pracovisko, počíta sa k UKF inak k externým, pre poriadok treba mať aj zoznam externých
* Pripájam nie celkom dokonalý kód, ktorý som používal v jave na rozbitie tohto záznamu – treba urobiť lepší v php (***src.rar***)
* Publikáciu rozbíjame do podoby minimálne:
  + autori: meno, priezvisko, pracovisko – toto je databáza autorov, nie zamestnancov, k tomu je možné pridať id\_zamestnanca, kde bude prepojenie na tabuľku zamestnancov



* + publikácie: crepc\_id je referenčné číslo z textovej podoby, k tomu kategória – zo záznamu hore je to V1, potom názov – po /, ale pozor niekedy sa / vyskytuje aj v názve. Autorov sem neukladáte, ale nasleduje popis, ktorý môžete rozbiť aj detailnejšie, ale nemusíte, nakoniec rok vydania. Rok vykazovania je k dispozícii až v ďalšom súbore.



* + autorstvo: rozbitie autorov z časti za „/“ do znaku „.“ Ak ten nie je súčasťou názvu ☹, pozor na typ autora v zásade nás zaujíma autor a ak nie je autor tak zostavovateľ a percentá pri nich:
  + 
  + ako zamestnanec\_id by u vás nemalo byť id z tabuľky autorov, ale id zamestnanca z tabuľky zamestnancov, ktoré je unifikované na celú univerzitu
* Všetky chyby, ku ktorý dôjde počas importu, je potrebné identifikovať a zobraziť v rozhraní, ktoré umožní úpravu na úrovni vstupu – navrhnite. Tieto údaje sa ďalej spracujú tak, aby nenastala nejaká duplicita, čiže pôvodný obsah prináležiaci tomuto záznamu vymažú a vložia sa nanovo – vhodné by bolo tým pádom vidieť aj zdroj (pôvodné údaje) ku každej publikácii – napr. cez prepojenie na id\_zdroja.

### Import akceptovaných publikácií

* toto je zoznam z ministerstva, kde sa nachádza zoznam publikácií identifikovaných len na základe národného registra CREPC, ktorý treba spárovať s publikáciami z predošlého importu
* xls/cvs súbor, v ktorom najdôležitejšiu časť predstavuje crepc\_id, na základe ktorého sa údaje párujú
* podoba sledovaných údajov je v súbore ***24957-fulldata.csv*** – dobre by bolo, aby ste umožnili definovať a vybrať stĺpce, ktoré sa majú importovať, stĺpce pre vašu kostru tabuľky zoberte z tohto súboru
* k údajom pridáte stĺpec, v ktorom bude aktuálny dátum, kedy import prebehol. Je to kvôli tomu, že asi tretina publikácií sa medziročne opakuje a mohlo sa zmeniť napr. ich zaradenie. Do starých dát nezasahujeme, pretože už na základe nich boli pridelené financie (nesmieme meniť parametre publikácie), ale v aktuálnom roku počítame so zmenenými. Pozor, vyplýva z toho, že crepc\_id môže byť v zozname 2x.

### Krížová kontrola

Je potrebné skontrolovať, či nejaké publikácie chýbajú alebo prevyšujú:

* Porovnáte údaje importu z detailným zoznamom publikácií a zoznamom z CREPC – výsledkom bude nejaká tabuľka, ktorá bude mať časť stĺpcov prázdnu na pravej strane a časť na ľavej.
* K tým, ktoré sú v CREPC a chýbajú v detailnom zozname umožníte ručne pridať detaily publikácie – autori, podiely atď. Zo zadaného obsahu sa následne vygenerujú správne záznamy.

## Párovanie

Po tomto celom divadle bude potrebné navrhnúť ešte spárovanie pracovísk z tohto systému a pracovísk z databázy zamestnancov.

Zmeny sa dejú vždy, takže tabuľka na spárovanie bude samostatná tabuľku pre dátum rovnaký ako má import do systému. Bude obsahovať pracovisko z ďalej uvedenej tabuľky zamestnancov a k nemu zoznam skratiek, ktoré sa majú do nich započítať.

Databáza zamestnancov má nasledovnú podobu, zobrazil som vám len podstatné polia.



Pre poriadok, tu je štruktúra:

**CREATE TABLE IF NOT EXISTS `zoznam\_osob\_ldap` (**

**`ldap\_id` int(11) NOT NULL,**

**`hash\_rodneho\_cisla` varchar(45) COLLATE utf8\_slovak\_ci NOT NULL,**

**`meno` varchar(45) COLLATE utf8\_slovak\_ci NOT NULL,**

**`priezvisko` varchar(100) COLLATE utf8\_slovak\_ci NOT NULL,**

**`rodne\_meno` varchar(255) COLLATE utf8\_slovak\_ci NOT NULL,**

**`plne\_meno` varchar(255) COLLATE utf8\_slovak\_ci NOT NULL,**

**`katedra` varchar(255) COLLATE utf8\_slovak\_ci NOT NULL,**

**`katedra\_kod` int(11) NOT NULL,**

**`katedra\_skratka` varchar(10) COLLATE utf8\_slovak\_ci NOT NULL,**

**`fakulta` varchar(255) COLLATE utf8\_slovak\_ci NOT NULL,**

**`fakulta\_kod` int(11) NOT NULL,**

**`aktivny` tinyint(4) NOT NULL,**

**`uvazok` float(5,2) NOT NULL,**

**`typ\_zam` varchar(50) COLLATE utf8\_slovak\_ci NOT NULL,**

**`okr\_zam` varchar(15) COLLATE utf8\_slovak\_ci NOT NULL,**

**`okr\_zam\_popis` varchar(255) COLLATE utf8\_slovak\_ci NOT NULL,**

**`okr\_zam\_popis\_ascii` varchar(255) COLLATE utf8\_slovak\_ci NOT NULL,**

**`kzam\_kod` varchar(15) COLLATE utf8\_slovak\_ci NOT NULL,**

**`kzam\_popis` varchar(255) COLLATE utf8\_slovak\_ci NOT NULL,**

**`kzam\_popis\_ascii` varchar(255) COLLATE utf8\_slovak\_ci NOT NULL,**

**`epca` int(11) NOT NULL,**

**`uid` varchar(50) COLLATE utf8\_slovak\_ci NOT NULL,**

**`uroven` int(11) NOT NULL DEFAULT '1',**

**PRIMARY KEY (`ldap\_id`),**

**KEY `hash\_rodneho\_cisla` (`hash\_rodneho\_cisla`),**

**KEY `aktivny` (`aktivny`),**

**KEY `katedra\_kod` (`katedra\_kod`),**

**KEY `katedra\_skratka` (`katedra\_skratka`),**

**KEY `uid` (`uid`)**

**) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8\_slovak\_ci;**

### View

Následne by malo byť možné zobrazovanie údajov zamestnanca. Údaje sa čerpajú z detailnej tabuľky a zobrazujú sa na základe prihláseného zamestnanca. Na prihlasovanie môžete vytvoriť tabuľku users s menom, heslom a ldap\_id za dodržania štandardných bezpečnostných prvkov.

## Projekty

Import projektov opäť spočíva vo viacerých krokoch:

### Import z databázy UKF

Z tohto dotazu sú zrejmé polia s údajmi, ktoré prídu do systému, predpokladajme csv (podoba je v prílohe ***employees\_projects2021.csv***)

**SELECT employees\_projects.employee\_id, employees\_projects.participation\_percentage, year(employees\_projects.date\_from), employees\_projects.project\_id, projects.title, projects.grant\_program\_id,**

**grant\_programs.acronym,**

**employees.priezvisko, employees.katedra, employees.fakulta,**

**project\_budgets.full\_expenditures**

**from employees\_projects**

**inner join projects on (projects.id = employees\_projects.project\_id)**

**inner join employees on (employees.id = employees\_projects.employee\_id)**

**inner join grant\_programs on (grant\_programs.id = projects.grant\_program\_id)**

**inner join project\_budgets on (projects.id = project\_budgets.project\_id)**

**where project\_budgets.\_year in (2020,2021)**

Skúste rozbiť na základe dotazu údaje na tabuľky a navrhnúť schému na pokrytie týchto potrieb. Položka ***employee\_id*** pochádza z univerzálnej tabuľky ***zoznam\_osob\_ldap***.

Údaje, ktoré sa budú importovať z tohto súboru by sa mali overovať – ak existujú údaje naviazané na rovnaký projekt, rovnakého riešiteľa a rovnaký rok, treba aby sa zaktualizoval podiel riešiteľa v danom roku, a prípadne rozpočet – tieto informácie by sa mali zobraziť aj prostredníctvom nejakého logu, že boli menené.

Zoznam zamestnancov je stále rovnaký

#### View

Z týchto údajov by malo byť možné zobraziť pre každého zamestnanca zoznam jeho projektov po rokoch a jeho podiel na celkovom rozpočte. Následne vo finále môžeme urobiť aj pekné zobrazenie, aktuálne stačí toto.

## Import z databázy ministerstva

Načítanie projektov z databázy ministerstva (excel s viacerými hárkami, zrejme rozdelíme na T1-T5 a pokojne môžu byt csv). Podobu údajov a v zásade aj štruktúru databázy máte v súbore Prehlad\_VVSprojekty\_2021.xlsx, pričom uvažujme len hárky T1-T5. Ideálne ak sa to pomestí do jednej spoločnej tabuľky a len v jednom stĺpci bude hodnota T1-T5, aby sme odlíšili kategórie. Treba dať pozor, aby sa importovali len záznamy UKF – umožníte vybrať, ktoré vyfiltrované záznamy importovať.

#### Párovanie

Následne pre tieto naimportované projekty je potrebné ich spárovanie s projektami v systéme k id\_čkam, ktoré sme naimportovali z databázy UKF. Moja predstava je zoznam ministerských projektov a prehľadávacie políčko vyhľadávajúce podľa názvu. Do ministerského importu za zapíše id projektu z databázy UKF. Súčasne sa môže skontrolovať, či sedia sumy v týchto projektoch.

## Import čiastkových databáz

Okrem tejto databázy je potrebné spárovať aj rezortné projekty, t. j. skontrolovať, či sa projekty z týchto databáz zhodujú s projektami v databázach UKF – niekedy sa na projekt zabudne a niekedy sa projekt len tak vymyslí.

Údaje sú k dispozícii v troch podobách: ***APVV***, ***VEGA***, ***KEGA***.

Prebehne import a kontrola podľa názvu, roku a pridelenej sumy - ak sa niečo líši, zobrazí sa a umožní úprava.