

# Countries in brief

## Zadání

Varianta B - Mashup nad LinkedData.

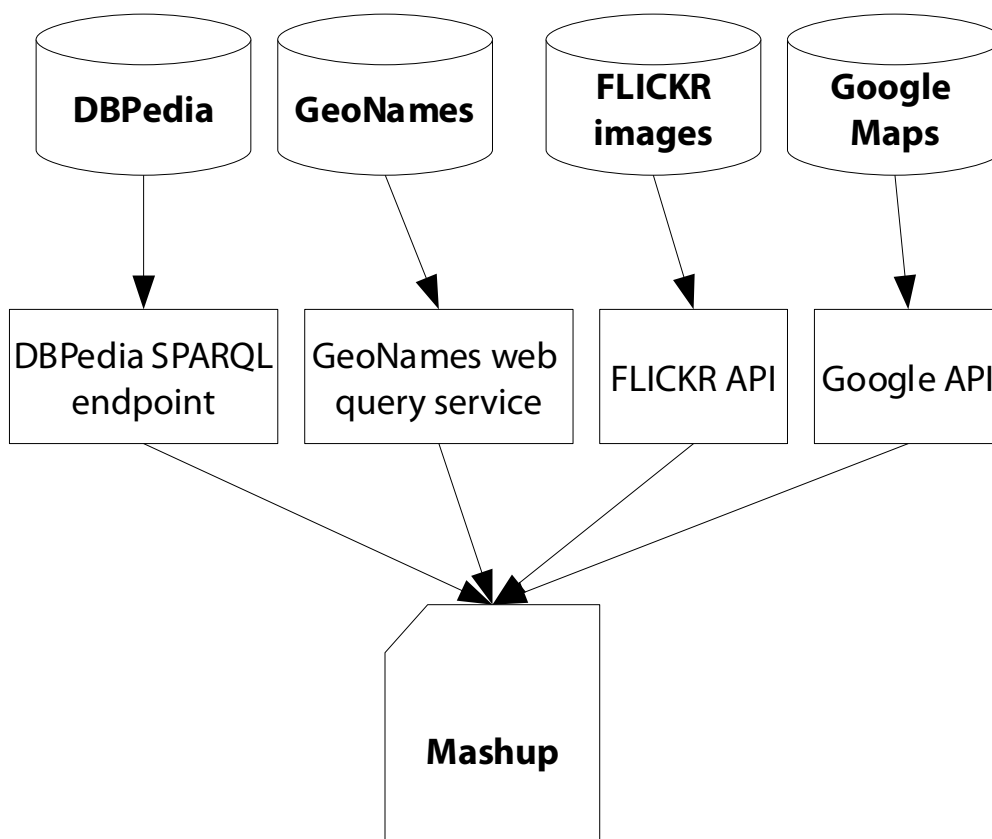
Vytvořte jednoduchý mashup, který bude využívat data minimálně ze dvou "linked datasetů" (viz. Linked Open Data cloud). Doporučené nástroje: PHP RDF API ARC2, Exhibit

## Úvod

Mashup s názvem „Countries in brief“ tvoří webová stránka, která umožňuje na základě vstupu, kterým je název nějakého státu na světě, zobrazit informace o daném státu, najít ho na mapě a vyhledat nějaké fotky, které se s daným státem pojí.

## Datové zdroje

Jako datová základna aplikace byly zvoleny 4 zdroje, z čehož 2 patří mezi Linked Data. Prvním a zároveň hlavním zdrojem dat je DBPedia. Tato rozsáhlá báze dat poskytuje informace o státech a odkazy na stránky wikipedie. Seznam států světa je získáván z GeoNames, zároveň s ním je získán i název hlavního města a geografické informace. Jako doprovodný vizuální materiál jsou použity fotky získané přes FLICKR API a mapa světa z Google maps API.



## Popis aplikace

Všechna funkcionalita aplikace je implementována na front-endu. Architektura je tedy postavena na HTML+CSS a JavaScriptu. Komunikaci se servery zajišťují AJAX požadavky, které realizuje javascriptová knihovna jQuery. Data jsou přijímána ve formátu JSON a dále jsou pak zpracována na výstup.

Vstupem pro aplikaci je jméno konkrétního státu. Pro usnadnění jeho zadávání je nejprve načten seznam států z GeoNames a pomocí metody autocomplete je nabídnut při zadávání. Z jména státu je vytvořen SPARQL dotaz do DBPedi, vzhledem k občasné nekonsistenci dat je připraven ještě jeden dotaz, který se při neúspěšném prvním pokusu pokusí data najít opakovaně.

První dotaz vyhledá všechny instance třídy Country (dbpedia-owl:Country). Nevýhodou je, že se nejedná jen o současné státy světa, ale o veškeré jejich historické varianty a další zaniklá impéria. Tudiž je nutné provést filtraci. Na základě průzkumu jednotlivých instancí bylo zjištěno, že státy, které již ukončili svou existenci mají atribut dbpedia-owl:dissolutionYear, nicméně ne vždy. Proto je jako pojistka ještě využito atributu dbpprop:yearEnd. Vyhledáme tedy pouze státy, které tyto atributy nemají.

```
SELECT ?comment ?wiki ?homepage
WHERE {
  ?uri rdf:type dbpedia-owl:Country.

  ?uri foaf:name ?name.
  FILTER regex(?name, currentCountry, "i")

  ?uri rdfs:comment ?comment.
  FILTER langMatches( lang(?comment), 'en')

  OPTIONAL { ?uri foaf:page ?wiki. }
  OPTIONAL { ?uri foaf:homepage ?homepage. }
  OPTIONAL { ?uri dbpprop:yearEnd ?year. }.
  OPTIONAL { ?uri dbpedia-owl:dissolutionYear ?year2. }.
  FILTER (!bound(?year))
  FILTER (!bound(?year2))
};
```

Při testování bylo zjištěno, že tento dotaz ve skutečnosti neposkytuje data pro všechny státy a autonomní oblasti s vlastní samosprávou, bylo tedy nutno přidat ještě jeden dotaz, který v případě neúspěchu prvního zbylé státy dohledal. V dotazu je využito jako hlavního parametru název hlavního města, neboť to je unikátní termín.

```
SELECT * WHERE {" +
  ?uri dbpedia-owl:capital ?capital." +
  ?capital rdfs:label capital@en." +
  ?uri rdfs:label ?label." +
  ?uri rdfs:comment ?comment." +

  OPTIONAL { ?uri foaf:page ?wiki. }
  OPTIONAL { ?uri foaf:homepage ?homepage. }
  FILTER langMatches( lang(?comment), 'en')
};
```

Po načtení informací o daném státu je podle geografických souřadnic z GeoNames nalezena poloha státu a přes Google API je načtena fyzická mapa. Současně je poslán požadavek na FLICKR API, který získá 10 fotek, které se té země týkají. Byla testována i varianta využití aplikace Flickrwrapp, která každému konceptu přiřazuje kolekci fotek, které ho vyjadřují. Přestože výsledky z Flickrwrapp jsou dobré, pro účely této aplikace by bylo dotazování na DBpedia značně redundantní, proto byla zvolena přímá varianta přes FLICKR API. Nicméně poznatky z testování Flickrwrappu byly využity.

## Výsledek

Na aplikaci je demonstrováno využití Linked Data na reálné webové stránce. Při implementaci nastalo mnoho problémů, zejména při tvorbě dotazů do DBPedi. Přestože pro testování dotazů je k

dispozici SPARQL Explorer, ve kterém je možné sestavit libovolný dotaz, nebylo snadné ho optimalizovat. Je to dáno občasnou nekonsistencí dat mezi GeoNames a DBpedia. Dalším velkým problémem je rychlost odezvy serverů. Jelikož je aplikace plně závislá na datech třetí strany, je tento nedostatek dost zásadní. V určitých chvílích je rychlost odezvy dbpedia.org v řádek desítek vteřin.

Aplikace je dostupná na adrese <http://martinstanek.cz/mashup>, je zatím otestována pouze ve Firefoxu, ostatní prohlížeče je teprve třeba odladit. Níže přikládám náhledy aplikace.

