



„Podporujeme výskumné aktivity na Slovensku/Projekt je spolufinancovaný zo zdrojov EÚ

Prehľad použiteľných zdrojov dát. Verejné dáta.

Predmetom tohto pokračovania druhej analýzy v rámci projektu bolo jeho doplnenie v oblasti dátového usmernenia. Táto správa je len formálnym výstupom analýzy vhodnosti rôznych verejných zdrojov dát, vybraných na základe ich prítomnosti v dátových štruktúrach systému súčasného sledovania pohybu výroby a spotreby elektrickej energie na Slovensku. Cieľom analýzy bolo nájsť a popísať ďalšie verejné zdroje dát vhodné k tvorbe inteligentných systémov, podporujúcich alebo nahradzujúcich dátové zdroje súčasného bilančného modelu estimácie odchýlky výroby a spotreby.

Pri tom sme prihliadali:

- na hodnotu metadát získaných z potenciálnych dátových zdrojov,
- na relevantnú legislatívu, najmä: 1.) Zákon č. 211/2000 Z. z. o slobodnom prístupe k informáciám a o zmene a doplnení niektorých zákonov, 2.) Zákon č. 275/2006 Z. z. o informačných systémoch verejnej správy, 3.) Štandardy pre informačné systémy úradov verejnej správy, 4.) Zákon č. 540/2001 Z. z. o štátnej štatistike a 5. ustanovenia na úrovni Európskej únie a medzinárodného práva,
- na bezpečnosť jednotlivých verejných zdrojov dát z pohľadu anonymity dát,
- na kvalitu potenciálnych dátových zdrojov,
- na náročnosť extrakcie a implementácie dát z verejných zdrojov do aktuálne bežiacich systémov,
- na relevanciu potenciálnych dátových zdrojov pri cieľoch krátkodobej predikcie (hodinový, denný interval) a cieľoch dlhodobej predikcie (týždenný a mesačný interval) spotreby a výroby elektrickej energie,
- na charakteristiky súčasného systému poskytovania, evidencie a využitia údajov o výrobe a spotrebe elektrickej energie.

Verejné dáta predstavujú zaujímavý zdroj poznania pre projekty oblasti výskumu a vývoja. Množstvo verejných zdrojov aj kvalita a objem dát v nich v posledných rokoch zaznamenávajú prudký nárast. Viaceré štáty, na čele s Veľkou Britániou

uvoľňujú štátne dáta a to nielen v agregovanej, ale aj surovej podobe. Ich motivácia na takúto aktivitu je dvojaká:

- a) občianske združenia a spolky sa dovolávajú svojich práv na prístup k informáciám, ktoré boli zbierané za ich dane a pokiaľ sa netýkajú bezpečnosti štátu, mali by byť zverejnené,
- b) zverejnenie dát umožňuje inováciu cez crowdsourcing mechanizmus.

Na nasledujúcich stranách popíšeme vybrané zdroje týchto dát vo svete, ktoré sa po našej analýze ukázali ako kvalitné a relevantné pre našu výskumnú činnosť.

Hydrometeorologické dáta

Francúzsko

Zdroj: <http://www.data.gouv.fr/en/dataset/donnees-d-observation-des-principales-stations-meteorologiques>

Popis zdroja: Na stránke sú k dispozícii hydrometeorologické dáta z územia Francúzska (nie len kontinentálneho), rozmiestnených v najväčších metropolách a geograficky relevantných bodoch.

Zbierané dáta: Zbierajú sa atmosférické parametre ako teplota, vlhkosť, smer a sila vetra, atmosférický tlak a hladina zrážok. Merané sú aj deskriptívne parametre ako časová senzitivita, vizibilita a meranie oblačnosti. Poslednou skupinou dát sú dáta špecifické k polohe stanice, napríklad hĺbka snehu a povrchový stav zeminy.

Kvalita dát: Na 63 staniciach (budú pribúdať) sa základné dáta zapisujú raz za 4 hodiny. Na serveri je v každom momente ponechaná dvojtyždňová história dát.

Náročnosť extrakcie a implementácie dát: Dáta sú voľne šíriteľné a stiahnuteľné v štandardnom formáte, ktorý umožňuje ich ľahkú manipuláciu. Dáta bude nutné zbierať pravidelne (raz za dva týždne) aby sa vytvorila relevantná vzorka. Túto aktivitu je možné zautomatizovať.

Hodnota získaných metadát: Prostredníctvom tohto zdroja máme k dispozícii dáta použiteľné na tréning algoritmov zapracovávajúcich hydrometeorologické trendy do celkovej predikcie.

USA

Zdroj: <http://www.nrel.gov/midc/>

Popis zdroja: Na stránke sú k dispozícii hydrometeorologické dáta z územia Spojených štátov amerických (nie len kontinentálneho), rozmiestnených v najväčších metropolách a geograficky relevantných bodoch.

Zbierané dáta: Dáta o osvite sú zbierané každou stanicou. Zoznam zbieraných premenných:

„Temperature (2m), Temperature (50m), Delta Temp (502m), Delta Temp (8050m), Dewpoint Temperature, Max Dewpoint Temperature, Time of Max Dewpoint Temperature, Min Dewpoint Temp, Time of Min Dewpoint Temp, Relative Humidity, Wind Speed (2m), Pk Wind Speed (2m), SDev Wind Speed (2m), Wind Direction (2m), Wind Direction at Pk WS (2m), SDev Wind Direction (2m), Time of Pk Gust (2m), Wind Speed (5m), Pk Wind Speed (5m), SDev Wind Speed (5m), Wind Direction (5m), Wind Direction at Pk WS (5m), SDev Wind Direction (5m), Time of Pk Gust (5m), Wind Speed (10m), Pk Wind Speed (10m), SDev Wind Speed (10m), Wind Direction (10m), Wind Direction at Pk WS (10m), SDev Wind Direction (10m), Time of Pk Gust (10m), Wind Speed (20m), Pk Wind Speed (20m), SDev Wind Speed (20m), Wind Direction (20m), Wind Direction at Pk WS (20m), SDev Wind Direction (20m), Time of Pk Gust (20m), Wind Speed (50m), Pk Wind Speed (50m), SDev Wind Speed (50m), Wind Direction (50m), Wind Direction at Pk WS (50m), SDev Wind Direction (50m), Time of Pk Gust (50m), Wind Speed (80m), Pk Wind Speed (80m), SDev Wind Speed (80m), Wind Direction (80m), Wind Direction at Pk WS (80m), SDev Wind Direction (80m), Time of Pk Gust (80m), Barometric Max, Time of Max Pressure, Min Pressure, Time of Min Pressure, Precipitation (Accumulated), Turbulence Intesity (2m), Turbulence Intesity (5m), Turbulence Intesity (10m), Turbulence Intesity (20m), Turbulence Intesity (50m), Turbulence Intesity (80m), Richardson Number (250m), Richardson Number (280m), Richardson Number (5080m), Friction Velocity (u), Est. Surface Roughness u* Quality Control (r2), Power Law Coef. (280m).“

Kvalita dát: Na 33 staniaciach sa základné dáta zapisujú s rôznou, ale dostatočnou frekvenciou. Tá variuje v závislosti od premennej od 10 minútových po denné merania.

Náročnosť extrakcie a implementácie dát:

Dáta sú voľne šíriteľné a stiahnuteľné v štandardnom formáte, ktorý umožňuje ich ľahkú manipuláciu. Datamining bude možné zautomatizovať.

Hodnota získaných metadát:

Tento unikátny zdroj nám prináša jedinečné dáta na tréning algoritmov zapracovávajúcich hydrometeorologické trendy do celkovej predikcie. Jeho výhodou je, že poskytuje aj predspracované vzťahy viacerých premenných, napríklad smer vetra pri jeho najväčšej a najmenej rýchllosti za jednotku času, trvanie doby a presný čas maxima a minima atmosférického tlaku, alebo štandardnú odchýlku rýchllosti a smeru vetra. Nevýhodou je, že ide o dáta z USA, takže ich môžeme použiť len na základný tréning a nemôžeme v nich hľadať špecifiká počasia nám relevantného (európske kontinentálne).

Tento zdroj môžeme doplniť o hydrometeorologické dáta z ďalšieho verejného zdroja z Hawaia: http://www.nrel.gov/midc/oahu_archive/ kde máme k dispozícii 1,5 ročnú históriu dát o osvite (W/m^2) z archívu systému Solar Measurement Grid. Dáta sú zapísané husto, na sekundovej báze. To nám umožňuje tréning systému na hľadanie trendov v osvite, ako jednej z hlavných hydrometeorologických premenných v energetike.

Hydrometeorologické dáta z verejných zdrojov sú v podobnej forme k dispozícii aj zo štátov Veľká Británia a Nórsko. Predpokladáme, že postupom času sa pridajú aj ďalšie štáty, táto téma je v súčasnosti v otvorenej diskusii aj u nemeckých zákonodarcov.

Energetické dáta

Veľká Británia

Zdroj: <http://www2.nationalgrid.com/uk/Industry-information/Electricity-transmission-operational-data/Data-Explorer/>

Popis zdroja: Prehľadové dáta o celkovom stave prenosovej sústavy v UK ponúkajú komplexný pohľad na sieť.

Zbierané dáta: Dátové rady o celkovom stave siete, prenosových aktivitách a výrobe solárnej a veternej energie.

Kvalita dát: Kvalitatívne primeraná frekvencia zapisovania dát. Databáza nám dáva k dispozícii dáta s 9 ročnou históriou (od 2005). Okrem toho publikuje aj predpoveď stavu siete na týždeň dopredu, fázovanú po dňoch.

Náročnosť extrakcie a implementácie dát: Dáta sú voľne šíriteľné a stiahnuteľné v štandardnom formáte, ktorý umožňuje ich ľahkú manipuláciu. Zber bude možné zautomatizovať.

Hodnota získaných metadát: Prostredníctvom tohto zdroja máme k dispozícii dáta ukazujúce vývoj a zaťaženie prenosovej sústavy na systémovej úrovni. Dáta sú preto použiteľné na tréning algoritmov v sledovaní trendov vo výrobe elektrickej energie. Chybovosť predpovedí, ktoré sa z týchto dát generujú na týždennej báze, nám môže slúžiť ako referencia výkonu nášho systému.

Zdroj: http://data.gov.uk/dataset/energy_consumption_in_the_uk

Popis zdroja: Zdroj predstavuje databázu údajov o spotrebe elektrickej energie a iných energií v UK.

Zbierané dáta: Štatistiky o spotrebe elektrickej energie sú tu rozdelené do piatich sekcií: celková spotreba energií, spotreba v doprave, domácnostiach, priemysle a v sektore služieb. Štatistiky obsahujú klasické premenné umiestnené v čase.

Kvalita dát: Zdroj drží dáta od roku 2007 do roku 2013. Ide o ročné súhrny spotrieb elektrickej energie podľa odvetví.

Náročnosť extrakcie a implementácie dát: Dáta sú voľne šíriteľné a stiahnuteľné v štandardnom formáte, ktorý umožňuje ich ľahkú manipuláciu. Zber bude možné zautomatizovať.

Hodnota získaných metadát: Tieto dáta nám umožnia pozrieť sa na vzťahy medzi spotrebou elektrickej energie a ostatných energií v rôznych ekonomických kontextoch a rôznom čase.

Zdroj: <http://data.gov.uk/dataset/energy-generation-from-solar-pv-arrays-for-selected-bristol-buildings>

Popis zdroja: Solárny smartmetering dataset z UK.

Zbierané dáta: Dáta sú z bežiacего projektu Energy UK z mesta Bristol. Zbierajú sa dáta o vygenerovanej elektrickej energii skupinami solárnych panelov (PV arrays), umiestnených na vybraných budovách v meste. Projekt má za cieľ testovať efektivitu zberu slnečnej energie v mestskej infraštruktúre.

Kvalita dát: Dáta sa zbierajú na 32 staniciach na dennej báze, hodnoty sú merané smartmetrami v krátkych intervaloch, generujú sa aj týždenné agregované dátové agregáty. História zberu závisí od stanice, najstaršie sú z roku 2012.

Náročnosť extrakcie a implementácie dát: Dáta sú voľne šíriteľné a stiahnuteľné v štandardnom formáte, ktorý umožňuje ich ľahkú manipuláciu. Zber bude možné zautomatizovať.

Hodnota získaných metadát: vzorka je zaujímavá, nakoľko je zbieraná na malom území (je tam vidno minimálne vzťahy) a smartmetringové dáta sú postupne agregované do použiteľnej dátovej štruktúry. Môžeme z nich odsledovať úroveň, do akej sa oplatí predprocesovať surové dáta zo smartmetrov ak by sme ich chceli použiť na analýzy rôznych úrovní, či ľahko interpretovateľnú vizualizáciu.

Nemecko

Zdroj:

<http://www.transparency.eex.com/en/Statutory%20Publication%20Requirements%20of%20the%20Transmission%20System%20Operators/Power%20generation/Actual%20wind%20power%20generation>

Popis zdroja: Stránka predstavuje zdroj dát o veterných a solárnych elektrárňach, primárne na území Nemecka.

Zbierané dáta: Zo zdroja sa vieme dozvedieť, koľko veternej a solárnej energie kto v Nemecku vyrobil. Sú tam aj dáta o výrobe solárnej a veternej energie na Slovensku, no nie je ich veľa.

Kvalita dát: Dáta zo staníc sa zbierajú a zapisujú v reálnom čase, výstupom sú denné agregované dáta o tom, koľko vyrobili jednotlivé bilančné skupiny, napríklad 50Hertz (DE), Amprion (DE), APG (AT), TenneT (DE), TransnetBW (DE). Historické dáta tu sú minimálne od začiatku roku 2010.

Náročnosť extrakcie a implementácie dát: Samotné dáta sú len ako ukážka z flashovej aplikácie. Predpokladáme, že so sťahovaním dát budú problémy, aj keď zdroj sa javí, že je verejným dátovým zdrojom.

Hodnota získaných metadát: Nemecká výroba solárnej a veternej energie má sprostredkovane relatívne veľký vplyv na výrobu na predikciu výroby elektrickej

energie na Slovensku. Zdroj by bolo dobré korelovať s nemeckými hydrometeorologickými dátami.

USA

Zdroj: <http://www.nrel.gov/midc/>

Popis zdroja: Vybrané dáta z „Clean energy“ produkcie na území Spojených štátov amerických.

Zbierané dáta: Na stránke sú k dispozícii dáta o zelenej produkcii, vybrané relevantne k USA hydrometeorologickým dátam z databázy popísaným v prvej časti reportu.

Kvalita dát: Dáta sú obmedzené na producentov, s geografickou blízkosťou ku staniciam, ktoré zbierajú hydrometeorologické dáta.

Náročnosť extrakcie a implementácie dát: Dáta sú voľne šíriteľné a stiahnuteľné v štandardnom formáte, ktorý umožňuje ich ľahkú manipuláciu. Zber bude možné automatizovať.

Hodnota získaných metadát: Zdroj umožňuje skúmať vzťahy medzi hydrometeorologickými trendami a výrobou.

Uvedené dáta je možné doplniť o databázu produkcie veterných elektrární v USA, ktoré zapisujú dáta v desať minútových intervaloch (variácia podľa jednotlivých premenných) zbieraných od roku 2004:

http://www.nrel.gov/electricity/transmission/western_wind_methodology.html

Nápomocné budú aj dáta zbierané z fotovoltaiických článkov v USA:

<http://wiki.dbpedia.org/Downloads39>

<https://openpv.nrel.gov/search>

Zaujímavá je aj databáza veterných turbín, popisujúca statické dáta o lokáciách a vlastnostiach (napríklad výkon) veterných jednotiek:

http://www.thewindpower.net/services_en.php#windfarms

Dáta o infraštruktúre a demografii

Zdroj: <http://data.worldbank.org/country>
<http://data.worldbank.org/country/slovakrepublic>

Popis zdroja: Dáta národnej banky o hospodárskych ukazovateľoch krajín (aj SR).

Zbierané dáta: Sú tam všetky relevantné dáta k analýze ekonomického rastu krajiny (aj SR).

Kvalita dát: Obsahuje 1300 ukazovateľov v číselných dátach za jednotlivé roky ku konkrétnej krajine, obsahujú aj dáta o elektrickej energii.

Náročnosť extrakcie a implementácie dát: Dáta sú voľne šíriteľné a stiahnuteľné v štandardnom formáte, ktorý umožňuje ich ľahkú manipuláciu. Zber bude možné zautomatizovať.

Hodnota získaných metadát: Cez tieto dáta vidíme trendy v spotrebe a smerovaní krajiny (aj SR). Je možné testovať vzťahy medzi ukazovateľmi ako export (mesačný, ročný) a spotrebou elektrickej energie. Vzorku možno doplniť o dáta o exporte produktov v čase:

http://www.thewindpower.net/country_zones_en_33_slovakia.php

Verejné zdroje demografických dát a dát o infraštruktúre:

<http://www.pmt.sk/Login.aspx?ReturnUrl=%2fstatistika.aspx>
<http://datahub.io/dataset/slovakian-company-registry>

Záver

Naším cieľom v tejto správe bolo urobiť prehľad použiteľných open-data zdrojov (verejné dáta), analyzovať ich legislatívne ukotvenie a ich kvalitu. Následne ohodnotiť možnosti ich využitia v systémoch, podporujúcich alebo nahradzujúcich dátové zdroje súčasného bilančného modelu estimácie odchýlky výroby a spotreby. Vymenované príklady verejných zdrojov dát a predovšetkým ich kombinácie majú potenciál posunúť vývoj algoritmov v slovenskej energetike na novú úroveň a zároveň reagovať na očakávané zmeny v kontexte dátovej štruktúry, ktorá bude analytickým subjektom k dispozícii v blízkej budúcnosti (smartmetering). Verejné dáta nie sú ešte dostatočne stabilné a bezpečné na to, aby mohli byť stálym zdrojom pre bežiaci finálny bilančný systém, no majú potenciál doplniť medzery vo vzorkách, použitých pri výskume a vývoji (tréningu) novej generácie algoritmov.

Kombinácia s dátami z neverejných zdrojov bude nutná, za účelom špecifikácie systému na slovenské prostredie. Nasledujúca správa sa bude venovať anonymizácii dát a dátových zdrojov, predpokladom a podmienkam, ktoré musia byť naplnené, metódam, nástrojom a automatizácii anonymizačných procesov. Návrhy na vylepšenie zberu verejných a neverejných dátových vzoriek, ich kvalita, popis metód a nástrojov ich anonymizácie, extrapolácie a vizualizácie budú predmetom samostatných správ.