



ÚSTAV PRO ČESKOU LITERATURU AV ČR, v. v. i.

✉ Na Florenci 3/1420, 110 00 Praha 1
🌐 www.ucl.cas.cz

☎ +420 224 818 437; fax +420 224 818 437
@ literatura@ucl.cas.cz

TISKOVÁ ZPRÁVA: Den s Korpusem českého verše

V Ústavu pro českou literaturu AV ČR byl dokončen unikátní projekt počítačnické versologie

V Praze 30. 9. 2014

Po několika letech intenzivní práce byl v Ústavu pro českou literaturu AV ČR, v. v. i., dokončen Korpus českého verše. **Korpus českého verše je svým rozsahem i zpracováním naprosto unikátní, a to i ve světovém měřítku. Díky on-line aplikacím a počítačovému programu, který umí automaticky rozpoznávat druh verše, získávají uživatelé dosud netušené možnosti práce s téměř 1 700 sbírkami české poezie 19. a počátku 20. století.** Všechny on-line aplikace jsou zdarma přístupné na stránkách Versologického týmu Ústavu pro českou literaturu AV ČR (www.versologie.cz).

V rámci Dne s Korpusem českého verše proběhne tisková konference, přednáška a křest publikace *Báseň a počítač*.

Tisková konference

Tisková konference proběhne **v úterý 4. 11. 2014 v přednáškovém sále ÚČL AV ČR od 10.30 hodin** (Na Florenci 1420/3, Praha 1, vchod C).

Přednáška

Přednáška se koná **v úterý 4. 11. 2014 v přednáškovém sále ÚČL AV ČR od 13 hodin** (Na Florenci 1420/3, Praha 1, vchod C).

Přednáška, která je součástí *Dne otevřených dveří ÚČL AV ČR* v rámci *Týdne vědy a techniky* (www.tydenvedy.cz), **představí** projekt automatické analýzy verše, tj. **výsledky počítačového programu**, který dokáže rozpoznat metrum českého verše. Přednášející **předvedou využití jednotlivých on-line aplikací** (Databáze českých meter, Eufonometr, Gunstick, Hex, Frekvenční slovníky, Cvičebnice). Posluchači se mimo jiné dozvědí, zda je v české poezii 19. století více jambů nebo trochejů, o čem je na základě frekvenční analýzy slov Máchův *Máj*, s čím vším rýmovali čeští básníci slovo láska nebo která česká báseň je nejefukoničtější.

Křest

Po přednášce bude od **14:30** následovat **křest publikace *Báseň a počítač*** (role kmotry se laskavě ujala PhDr. Květa Sgallová, CSc.), kterou vydává nakladatelství Academia v edici *Věda kolem nás* (<http://www.academia.cz/edice/veda-kolem-nas.html>).

Další informace

@ Mgr. Petr Plecháč, Ph.D.: plechac@ucl.cas.cz
@ PhDr. Robert Ibrahim, Ph.D.: ibrahim@ucl.cas.cz
🌐 www.versologie.cz

Báseň a počítač

Robert Ibrahim – Petr Plecháč

ÚČL AV ČR / Academia, Praha 2014

Anotace: Báseň a počítač? Jde takové spojení vůbec dohromady? Nestojí poezie a výpočetní technika v protikladu? Využití výpočetní techniky v přírodních a exaktních vědách nikoho nepřekvapí, chápe se jako samozřejmé. Jako samozřejmé se ale postupně začalo chápat také to, že počítače se staly nedílnou součástí našeho života, a pronikly tedy do všech oblastí, kterými se člověk zabývá. Užití výpočetní techniky v humanitních vědách má svou tradici – velmi plodně se rozvíjelo zejména v šedesátých letech 20. století, kdy počítače (tehdy obrovské mašiny, které umělo obsluhovat jen pár specialistů) pomáhaly zpracovávat velké datové soubory. Nikdy to ale nefungovalo tak, že by – a trochu přeháníme – literární vědec vytrhl báseň ze sbírky a podobně jako ve filmu *Jáchyme, hod' ho do stroje* (1974) ji „hodil“ do počítače, který za okamžik vyplivne výsledek. Literární vědec samozřejmě musel počítači data připravit – odměnou mu bylo, že počítač data setřídil a umožnil v nich pohodlně vyhledávat. Novou etapu ve využití počítačů znamenala devadesátá léta, období prudkého nástupu výpočetní techniky a jejího ovládnutí světa. Se zdokonalením počítačů a rozšířením jejich možností také rostlo jejich využití v humanitních vědách. Využití počítačů v literární vědě se váže k tomu, že objekt výzkumu této vědy, tj. literární texty, existuje v elektronické podobě. Vše, o čem je v publikaci řeč, by nikdy nevzniklo bez České elektronické knihovny (www.ceska-poezie.cz), která byla vytvořena v Ústavu pro českou literaturu AV ČR.

Publikace pojednává o tom, co všechno se dá s literárními texty za využití moderní výpočetní techniky dělat. Většina kapitol se týká verše, nejprve se tedy čtenář stručně seznámí s disciplínou, která se veršem zabývá (kapitola Co je to versologie?). V další kapitole je řeč o tom, co je potřeba udělat, aby z elektronických textů vznikl elektronický korpus, tj. takový soubor textů, který je připraven ke složitějšímu vyhledávání a k zadávání složitějších dotazů (kapitola Korpus českého verše). Poté se čtenář seznámí s nejrůznějšími aplikacemi, které umožňují pracovat s daty z korpusu: detailní informace o českém verši 19. století poskytuje Databáze českých meter (viz stejnojmennou kapitolu); která slova se v 19. století rýmovala se slovem „láska“, ukáže aplikace Gunstick (kapitola Láska : páska); odpověď na otázku, která česká báseň je nejefoničtější, podá aplikace Eufonometr (kapitola Dá se spočítat eufonie?); o čem je Máchův *Máj* a které podstatné jméno se v něm objevuje nejčastěji, prozradí aplikace Hex a Frekvenční slovníky (viz stejnojmenné kapitoly); a nakonec se možná dostaneme k tomu, jak zjistit, kdo napsal *Rukopisy*...



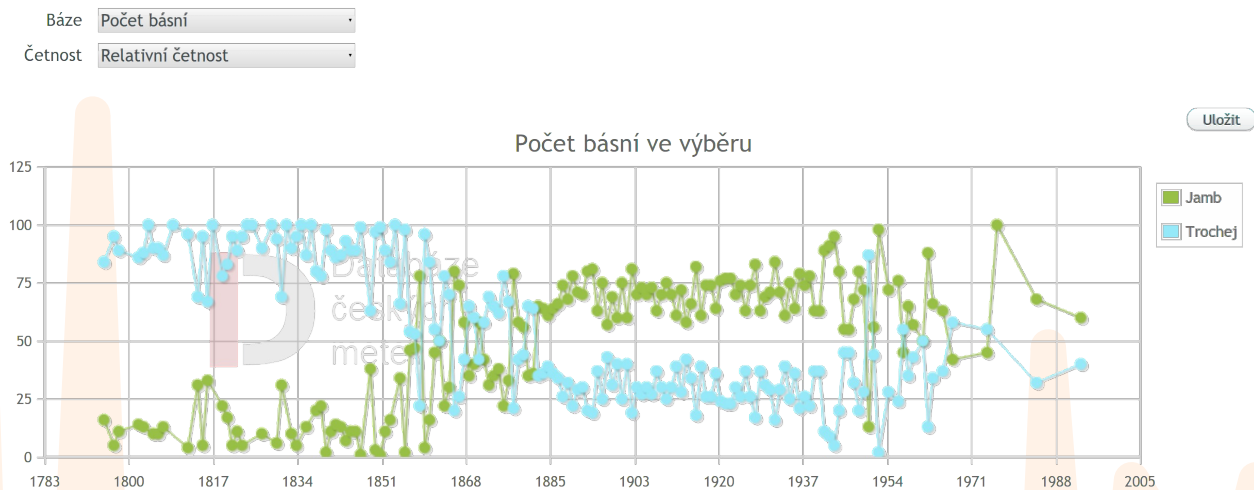
Stručné představení Korpusu českého verše a jeho on-line aplikací

Korpus českého verše (<http://versologie.cz/kcv.html>)

Korpus českého verše je lemmatizovaný, foneticky, morfologicky, metricky a stroficky anotovaný korpus české poezie 19. a počátku 20. století. Ke každé slovní jednotce v korpusu je připojena informace o jejím základním slovním tvaru (lemma), fonetickém přepisu a gramatických kategoriích, u každého verše je určeno metrum (jamb, trochej, ...), rozsah (čtyřstopý, pětistopý, ...), typ klauzule (mužská, ženská, ...) a metrický vzorec. Na vyšších rovinách jsou pak anotovány rýmové dvojice a pevné formy (sonet, rondel, ...).

Databáze českých meter (<http://versologie.cz/dcm/>)

V Databázi českých meter je zpřístupněn metrický a strofický popis 1 689 básnických sbírek (78 391 básní) z 19. a počátku 20. století obsažených v Korpusu českého verše. V Databázi lze (kromě standardních bibliografických údajů) vyhledávat a statisticky vyhodnocovat data dle veršového rozměru, metrického vzorce, rýmovanosti a strofičnosti textu, strofického a rýmového schématu, ustáleného názvu rozměru (alexandrin, hexamet, ...), pevných forem (sonet, rondel, ...).



Eufonometr (<http://versologie.cz/eufonometr/eufonometr.html>)

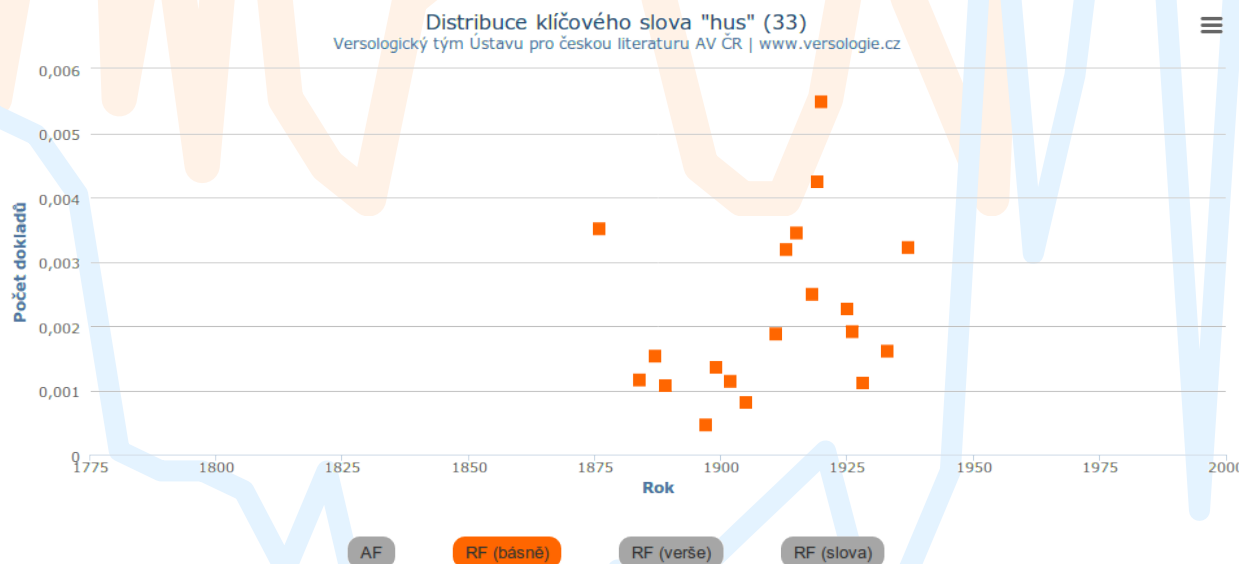
Aplikace Eufonometr umožňuje vypočítat eufonický koeficient libovolného textu, tj. nalézt takové opakování hlásek, které je statisticky významné. Uživatel do aplikace napíše nebo vloží text a zobrazí se mu jak eufonicky relevantní hlásky a jejich eufonický koeficient v jednotlivých řádcích, tak celkový eufonický koeficient vloženého textu a autor, jehož průměrný eufonický koeficient je nejbližší eufonickému koeficientu vloženého textu. Z dat zobrazených v grafu pak lze zjistit, kolik básní v České elektronické knihovně dosahuje obdobného eufonického koeficientu jako vložený text.

HEX - klíčová slova v české poezii (<http://versologie.cz/hex/>)

Aplikace Hex umožňuje vyhledávat v Korpusu českého verše texty, které obsahují uživatelem specifikované klíčové slovo, nebo naopak u uživatelem specifikovaného okruhu textů zobrazit všechna klíčová slova v nich nalezená.

DISTRIBUCE KLÍČOVÉHO SLOVA "HUS" (33) AUT: | SB: | - (A = 0.001; N ≥ 3; POS: NAV)

[NOVÉ HLEDÁNÍ](#) | [O APLIKACI](#)

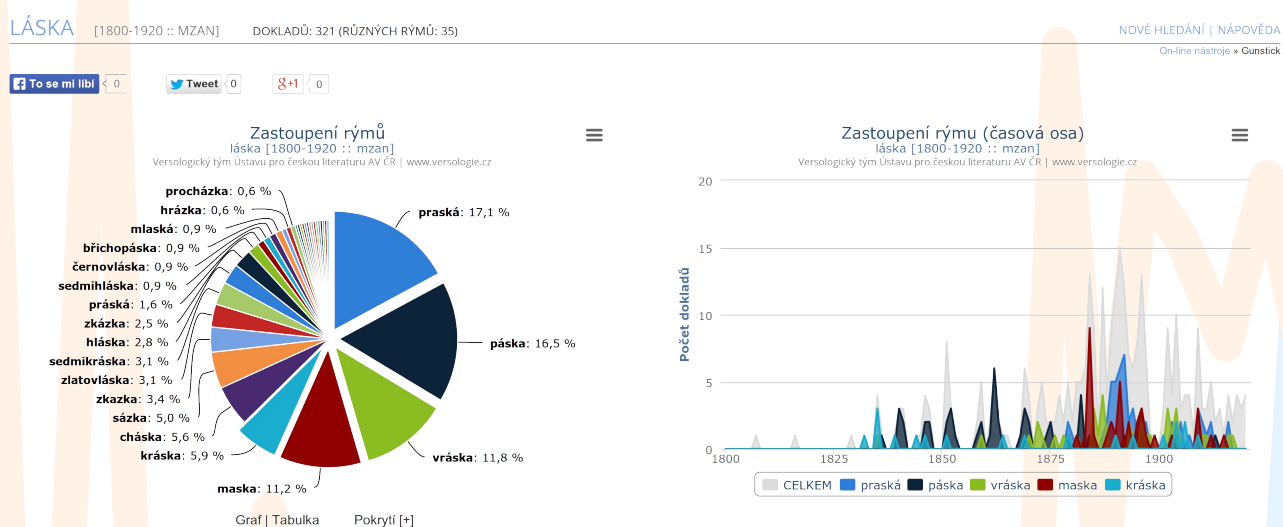


Frekvenční slovníky (<http://versologie.cz/slovníky/slovníky.html>)

Frekvenční slovníky české poezie obsahují údaje o četnosti slov v básnických dílech zahrnutých do Korpusu českého verše. Slovníky udávají jednak frekvenci lemmat, jednak frekvenci slovních tvarů (tokenů), a to jak v jednotlivých básnických sbírkách, tak v celých autorských subkorpusech i v celém Korpusu českého verše. Právě údaje o frekvenci lemmat odlišují tyto slovníky od Frekvenčních slovníků České elektronické knihovny.

Gunstick – databáze rýmů (<http://versologie.cz/gunstick/>)

Databáze českých rýmů zpřístupňuje výsledky automatické analýzy rýmových shod v básnických sbírkách obsažených v Korpusu českého verše a vydaných do roku 1920 (databáze obsahuje přes jeden milion rýmových párů). Tyto výsledky byly podrobeny ruční kontrole, která zohlednila různé kontexty (sbírky, básně, dobové poetiky...). Databáze obsahuje jak standardní rýmové shody, tak i některé případy volnějších shod.



Něco navíc

Cvičebnice (<http://versologie.cz/cvicebnice/>)

Interaktivní cvičebnice doplňuje publikaci *Úvod do teorie verše* (Ibrahim, Robert – Plecháč, Petr – Říha, Jakub; Akropolis, Praha 2013). Uživatel může procvičovat své znalosti při poznávání meter i celých rozměrů sylabotónických básní (více než 50 000 příkladů), pevných forem v sylabotónickém verši (více než 8 000 příkladů) a časoměrných útvarů (více než 300 příkladů). Každý uživatel má přehled o své úspěšnosti v jednotlivých kategoriích. Cvičebnici doprovází podrobný výklad.