

Petr Kadlec, Wikikonference 2011

# **METADATA NA WIKIPEDII**

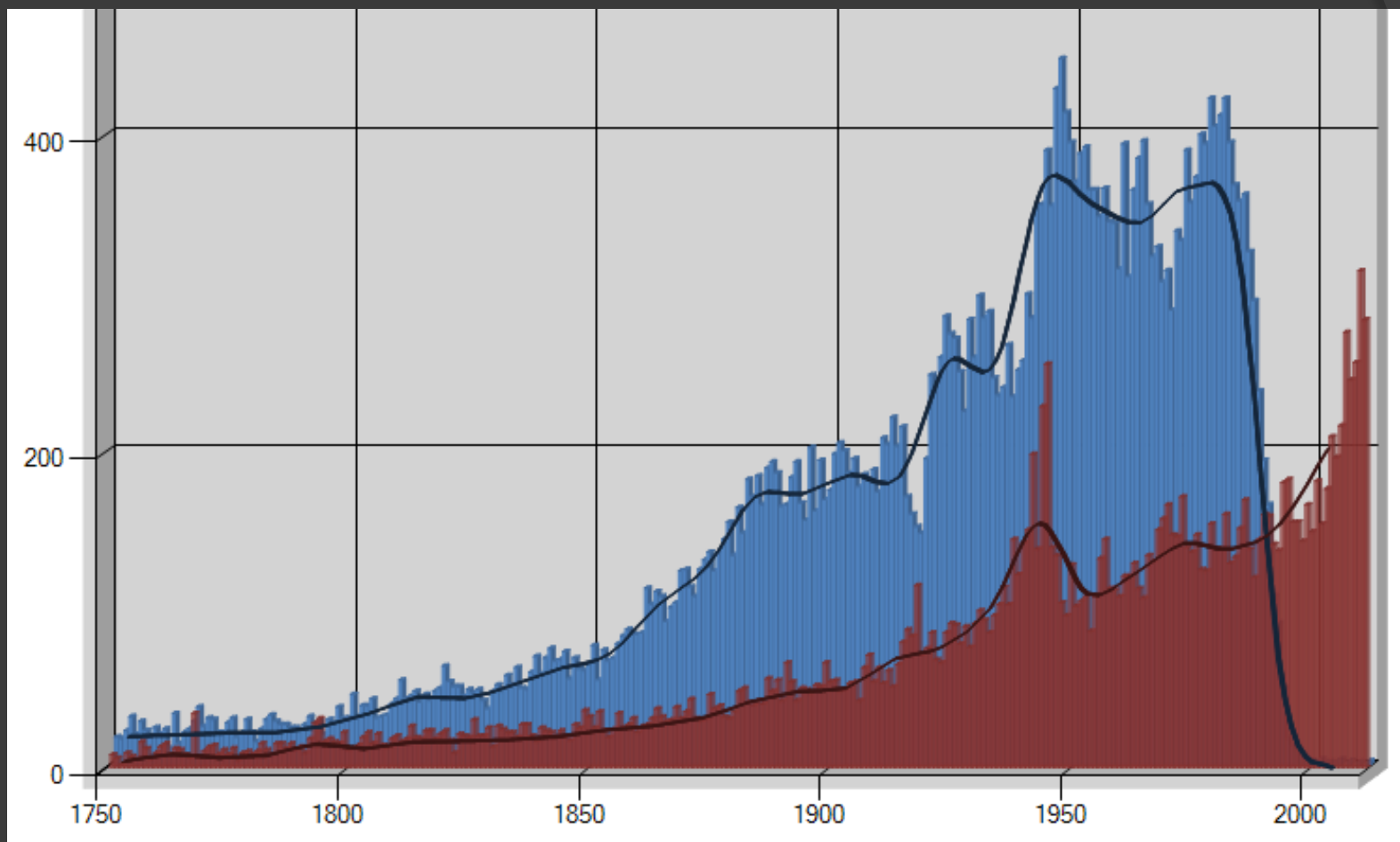
# Wikipedie

- ⦿ ...je velmi široký zdroj informací
  - Spisovatelé
  - Místa, budovy, ...
  - Chemické látky
  - Choroby
  - Léky
  - ...

# ...pro každého

- ⦿ Články obsahují velké množství různých informací
- ⦿ Pokud člověk umí česky, může se ledacos dozvědět...
  - Třeba u biografií rok narození a úmrtí příslušné osoby

...ale co když chci takový graf?



# Stroje neumějí číst česky

- ⦿ Z běžného textu v češtině se data pro strojové zpracování získávají těžko.
- ⦿ Je potřeba strojům pomoci.
  - Tenhle graf mohl vzniknout jen díky existenci specializovaných kategorií narození a úmrtí.

# Metadata

- ⦿ Jako „metadata“ se označují „data o datech“
- ⦿ Data (samotný článek) se vhodným způsobem doplní, „označují“ se v něm obsažené údaje

# Třeba... kategoriemi

- ⦿ Wikipedie má články řazené do kategorií
- ⦿ ...má svůj samodomo kategorizační „strom“
  - Je poněkud samorostlý
  - Není to strom
    - dokonce ani není acyklický
      - Debian ↔ Distribuce založené na Debianu
  - Ale určitě velmi užitečná věc!

# Děláme to, jak chceme

- ⦿ Wikipedie se encyklopedickým obsahem nesnaží nahradit či překonat „klasické“ zdroje
- ⦿ Je vyžadováno používání a uvádění zdrojů informací
- ⦿ Oproti tomu náš kategorizační strom se k žádným „autoritám“ nevztahuje



# Jak to dělají profesionálové

- ⦿ Kategorizační a podobné systémy byly už před Wikipedií:
  - Mezinárodní desetinné třídění
  - Třídění Kongresové knihovny
  - Medical Subject Headings
  - Polytematický strukturovaný heslář
  - ...

# Děláme to dobře?

- ⦿ Vlastní výzkum? NPOV?
- ⦿ Nedělají to jiní lépe?
  - Vždyť ani nemáme šanci porovnávat!
- ⦿ Dokáže odborník zvyklý na své kategorie používat ty naše?

# Jak srovnat?

- K tomu všemu potřebujeme nějaký způsob, jak provázat naše kategorie s jinými kategorizačními systémy

# Nejde ale jen o kategorie

- ⦿ Wikipedie je velmi široký zdroj informací
- ⦿ V žádné jednotlivé oblasti ale není sama
  - Autoritní záznamy v knihovnických databázích
  - GISy, mapy
  - Chemical Abstracts Service
  - Mezinárodní klasifikace nemocí
  - Anatomicko-terapeuticko-chemická klasifikace léčiv
  - ...

# Databáze autorit NK ČR

- NK ČR spravuje tzv. národní autority ČR

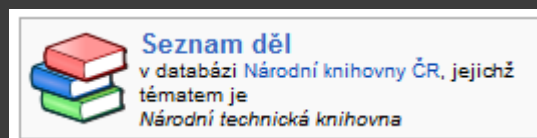
<b>Ident. číslo</b>	jk01021023
<b><u>Záhlaví</u></b>	● <a href="#">Čapek, Karel, 1890-1938</a>
<b><u>Pseudonym</u></b>	Plocek, Karel, 1890-1938
<b><u>Pseudonym</u></b>	Vašek, Karel, 1890-1938
<b><u>Odkaz. forma</u></b>	Capek, Karel, 1890-1938
<b><u>Odkaz. forma</u></b>	Chapekku, Kareru, 1890-1938
<b><u>Odkaz. forma</u></b>	Tchápek, Karel, 1890-1938
<b><u>Odkaz. forma</u></b>	Çapek, Karel, 1890-1938
<b><u>Odkaz. forma</u></b>	Čapekas, Karelas, 1890-1938
<b><u>Odkaz. forma</u></b>	Tšábík, Kār, 1890-1938
<b>Biograf. údaje</b>	Narozen 9.1.1890 v Malých Svatoňovicích u Trutnova, zemřel 25.12.1938 v Praze. PhDr., prozaik, žurnalista, dramatik, esejista, básník, autor knih pro děti, překladatel z francouzštiny, literární, výtvarný a divadelní kritik, estetik, filosof, filmový libretista.
<b>Status</b>	definitivní
<b>Zdroj</b>	Hordubal Literární noviny 46/2001, str. 16 - pseudonymy PNP-LA

# Authority NK ČR a Wikipedie

- ◉ Wikipedie v odpovídajících článcích obsahuje identifikátor onoho autoritního záznamu
- ◉ Vkládán ručně běžnou editací
- ◉ Šablony
  - `{{NK ČR}}`
  - `{{ID authority}}`

# To umožňuje

- Z článku odkázat na seznam děl v databázi NK ČR



- Z autoritního záznamu NK ČR (automaticky) odkázat na odpovídající článek Wikipedie

## Další informace



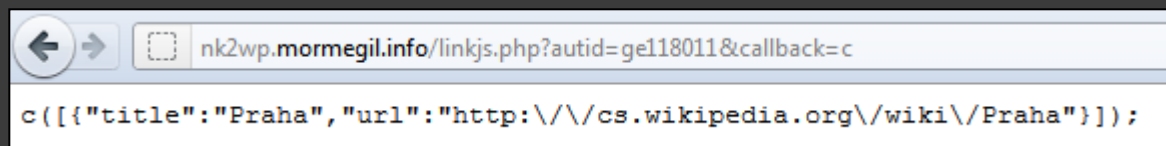
Oficiální stránka korporace: <http://www.techlib.cz/cs/>



[Wikipedie \(Národní technická knihovna\)](#)

# Anebo...

- Získat název odpovídajícího článku a odkaz ve snadno strojově zpracovatelné podobě *komukoli*, kdo zná onen identifikátor autoritního záznamu (třeba jiné knihovně)

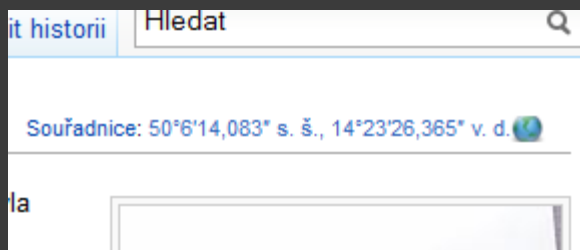


The image shows a browser window with the address bar containing the URL `nk2wp.mormegil.info/linkjs.php?autid=ge118011&callback=c`. Below the address bar, the page content displays a JSON object: `c([{"title":"Praha","url":"http:\\\\cs.wikipedia.org\\wiki\\Praha"}]);`



# Zeměpisné souřadnice

- Článek o místu typicky obsahuje souřadnice odkazující na mapy



- ...ale dokážete najít článek o místu, na kterém zrovna jste?

# Wikitude?



File:Wikitude.jpg z Wikimedia Commons, autor: Mr3641, dostupné pod licencí Creative Commons Uvedte autora-Zachovejte licenci 3.0 Unported

# Dosud spousta práce

- ⦿ Každý s podobným nápadem se musí prohrabat haldou (wiki)textu.
- ⦿ ...ale co kdyby Wikipedie poskytovala službu „dej mi seznam článků poblíž zadaných souřadnic“?

# K tomu je potřeba...

- Museli bychom některá „zajímavá“ data ve Wikipedii označovat nějakým standardizovaným způsobem

# WikiProjekt Metadata

- ⦿ Příprava základních nástrojů pro
  - Standardizované vkládání metadat do článků a kategorií
  - Jednoduché (a jednotné) využívání metadat vnějšími uživateli

# Příprava konkrétních aplikací

- ⦿ Do článků
  - NK ČR stačí převést
  - Souřadnice?
- ⦿ Vnější kategorizační systémy
  - ATC kategorie
  - PSH

# Jak to může vypadat?

## ☉ Kategorie

Hlavní článek: [Fotbal](#) • Portál: [Fotbal](#)

Sesterské projekty: [Wikimedia Commons](#) • [Wikislovník](#)

Externí kategorizační systémy: PSH: [fotbal](#) [↗](#) • LCSH: [Soccer](#) [↗](#) • MDT: [796.332 Fotbal \(kopaná\)](#) [↗](#)

## ☉ Metadata v článcích

Portály: [Informační věda a knihovnictví](#)

Metadata: [Mapy](#) | [Díla v NK ČR](#) | [Archiweb](#)

Kategorie: [ČVUT](#) | [Dejvice](#) | [Knihovny v Praze](#)

# Názory, nápady?

- ⦿ <http://wpcz.org/Wikipedista:Mormegil>
- ⦿ [mormegil@centrum.cz](mailto:mormegil@centrum.cz)