

# Současná lexikografie

PLIN035: Počítačová lexikografie

František Kovařík

**Opáčko**

**DWS**

**Opáčko**

DWS

**Markup**

Opáčko

DWS

Markup

**Data models**

Opáčko

DWS

**Grammars**

Markup

Data models

**Opáčko**

DWS

Grammars

Markup

**Přístup k datům**

Data models

**Opáčko**

DWS

Grammars

Markup

Přístup k datům

Data models

**Post-editing**

**lexicography**

# Dictionary Writing Systems



# Dictionary Writing Systems

... jako editační nástroj.

... jako managerský nástroj.

... jako publikační nástroj.

# DWS jako editační nástroj

Robustní.

Intuitivní.

Rychlý. Efektivní. Online.

# DWS jako managerský nástroj

Kontrola (editora/slovníku).

Distribuce dávek.

Trénink. Feedback.

# DWS jako publikační nástroj

Bezpečnost dat.

Tvorba výstupu.

Opakované použití dat.

# DWS

1. Front-end.

2. „Behind front-end“

(text v databázi).

# Datové struktury

# Document type definition (DTD)

## Document type definition (DTD)

Komponenty + jejich hierarchie.

Zajišťuje stejnou  
strukturu záznamů.



**Proč nepsat slovníky ve Wordu?**

**Proč specializovaný software?**

Snadná konverze dat.

Strukturovanost, uniformovanost.

Zobrazení nezávislé na obsahu.

```
<entry>
  <headword>bank</headword>
  <pos>noun</pos>
  <sense>
    <definition>institution you send your money to</definition>
    <example source="...">...</example>
    <example source="...">...</example>
    <translation>banka</translation>
  </sense>
  <sense>
    <definition>area along the side of a river</definition>
    <example source="...">...</example>
    <example source="...">...</example>
    <translation>břeh</translation>
  </sense>
</entry>
<entry>
...
```

XML

```
{ "entry": {
  "headword": "bank",
  "pos": "noun",
  "senses": [
    { "definition": "institution you send your money to",
      "examples": [
        { "source": "...",
          "text": "..."}], ...
    ],
    "translation": "banka"
  },
  { "definition": "area along the side of a river",
    "examples": [
      { "source": "...",
        "text": "..."}
    ],
    "translation": "banka"
  }
]
}
```

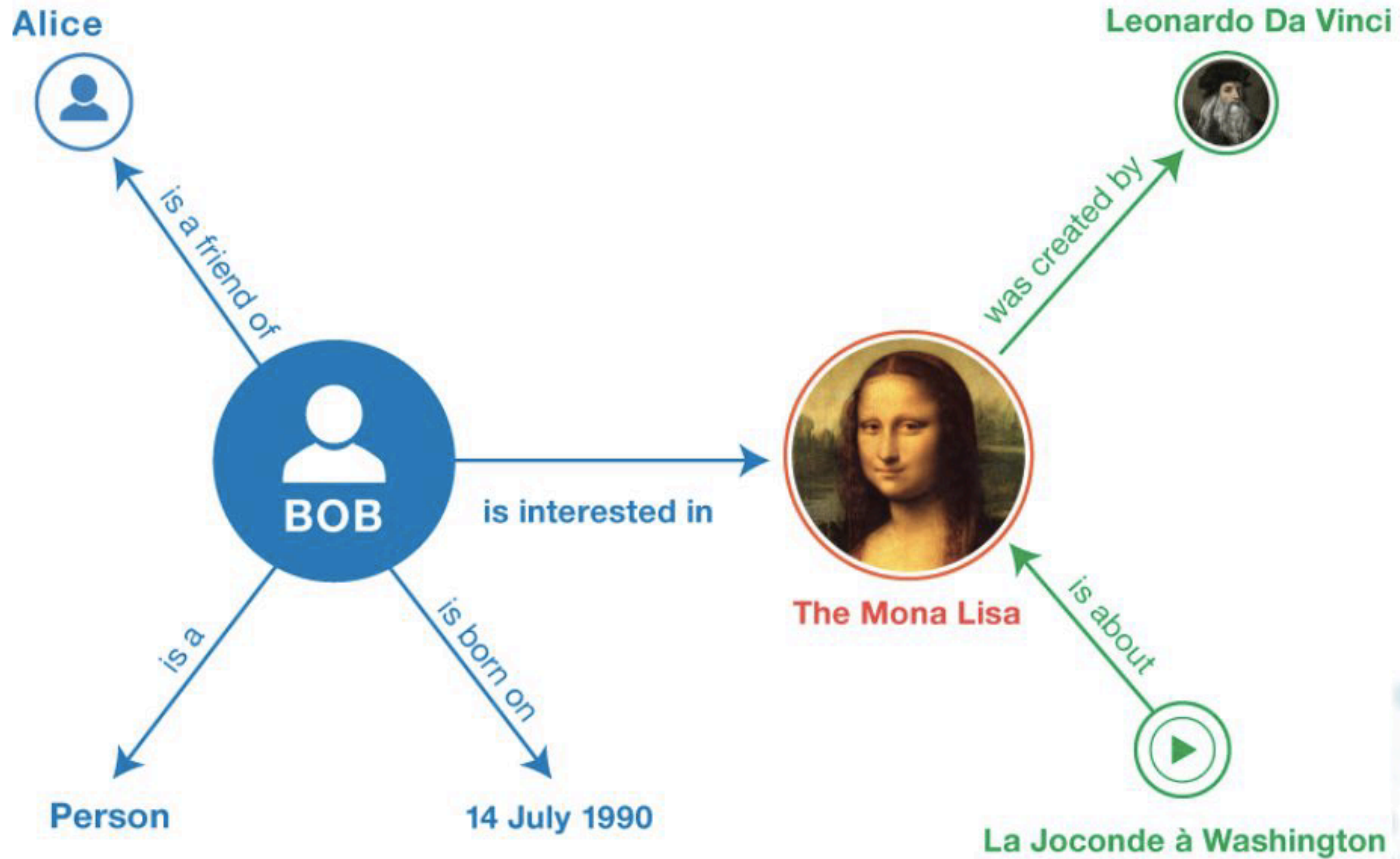
# JSON

**Zajímavé věci (nebudou na zkoušce)**

**Resource Description Network**

**Lemon: Lexicon Model for Ontologies**

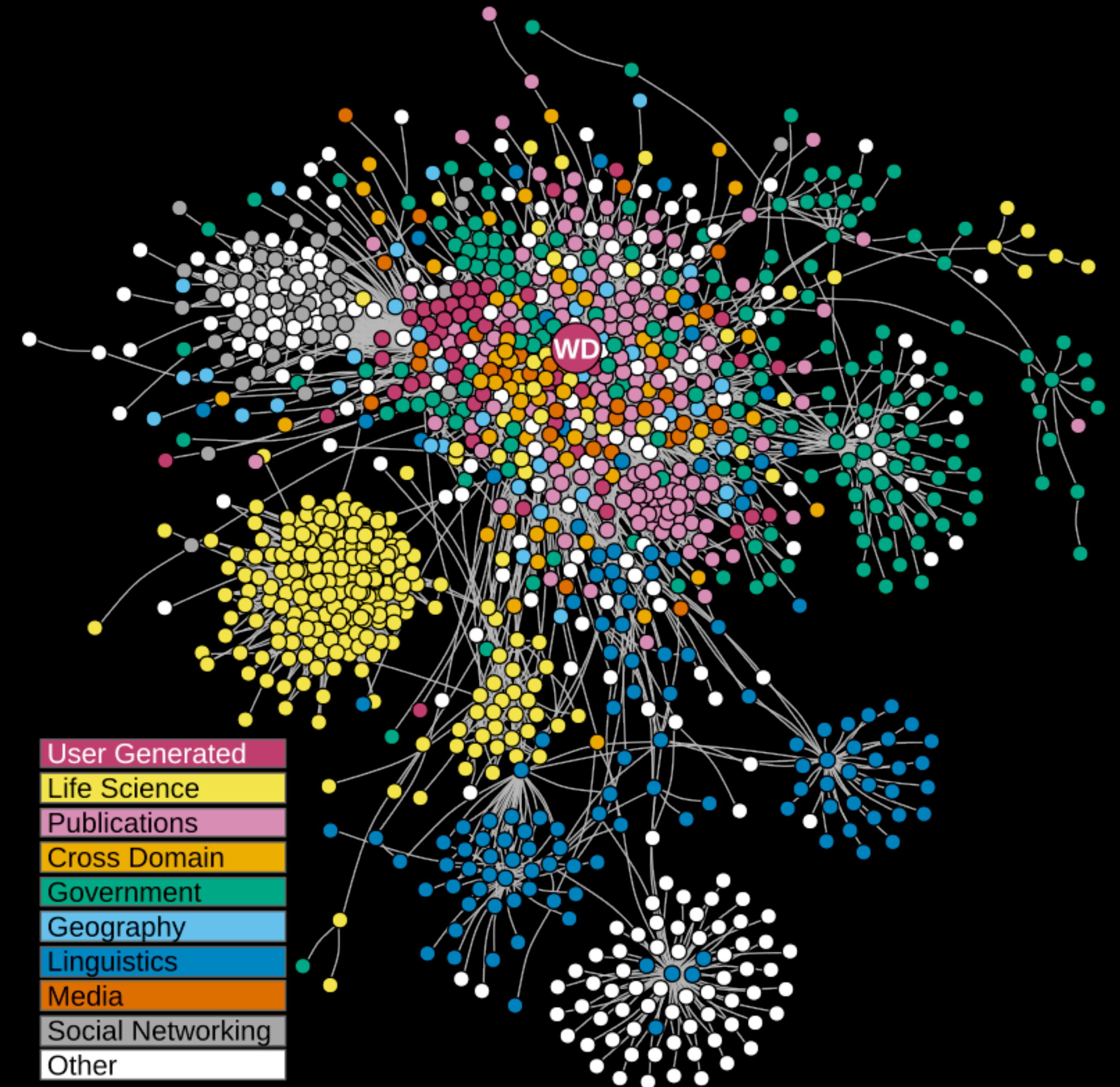
**Linked (Open) Data**



# Resource Description Network

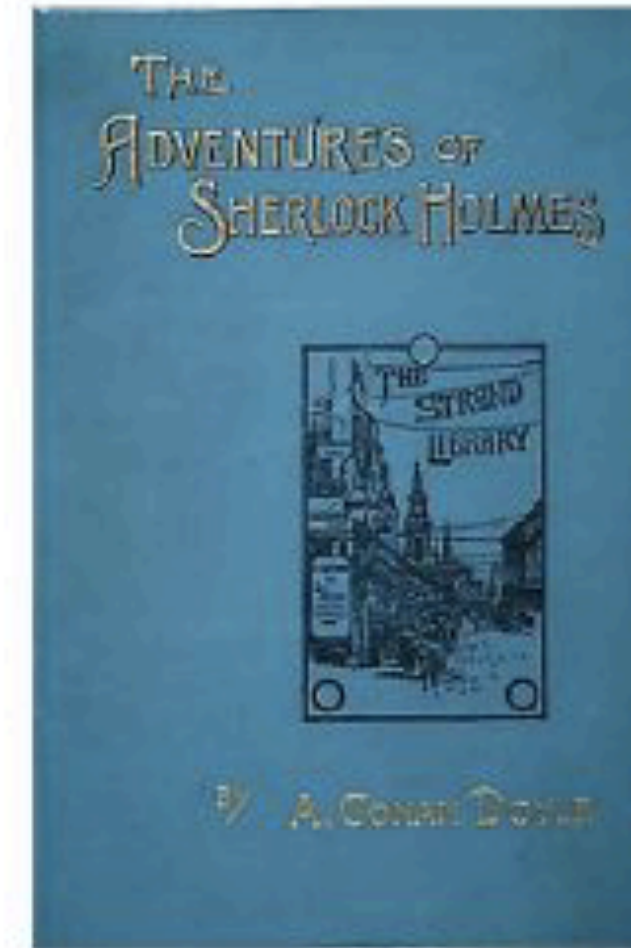
# Linked (Open) Data (LOD)

Princip webu.  
Čitelné počítačem.



## Knihy

```
<text>
  <front>
    <head rend="italic">Adventures of Sherlock Holmes</head>
    <docTitle>
      <titlePart>Adventure II. —</titlePart>
      <titlePart>The Red-Headed League</titlePart>
    </docTitle>
    <byline>By A. Conan Doyle.</byline>
  </front>
  <body>
    <p>I had called upon my friend, Mr. Sherlock Holmes, one day
      in the autumn of last year and found him in deep conversation
      with a very stout, florid-faced, elderly gentleman with fiery red hair ...
    </p>
  </body>
</text>
```



XML



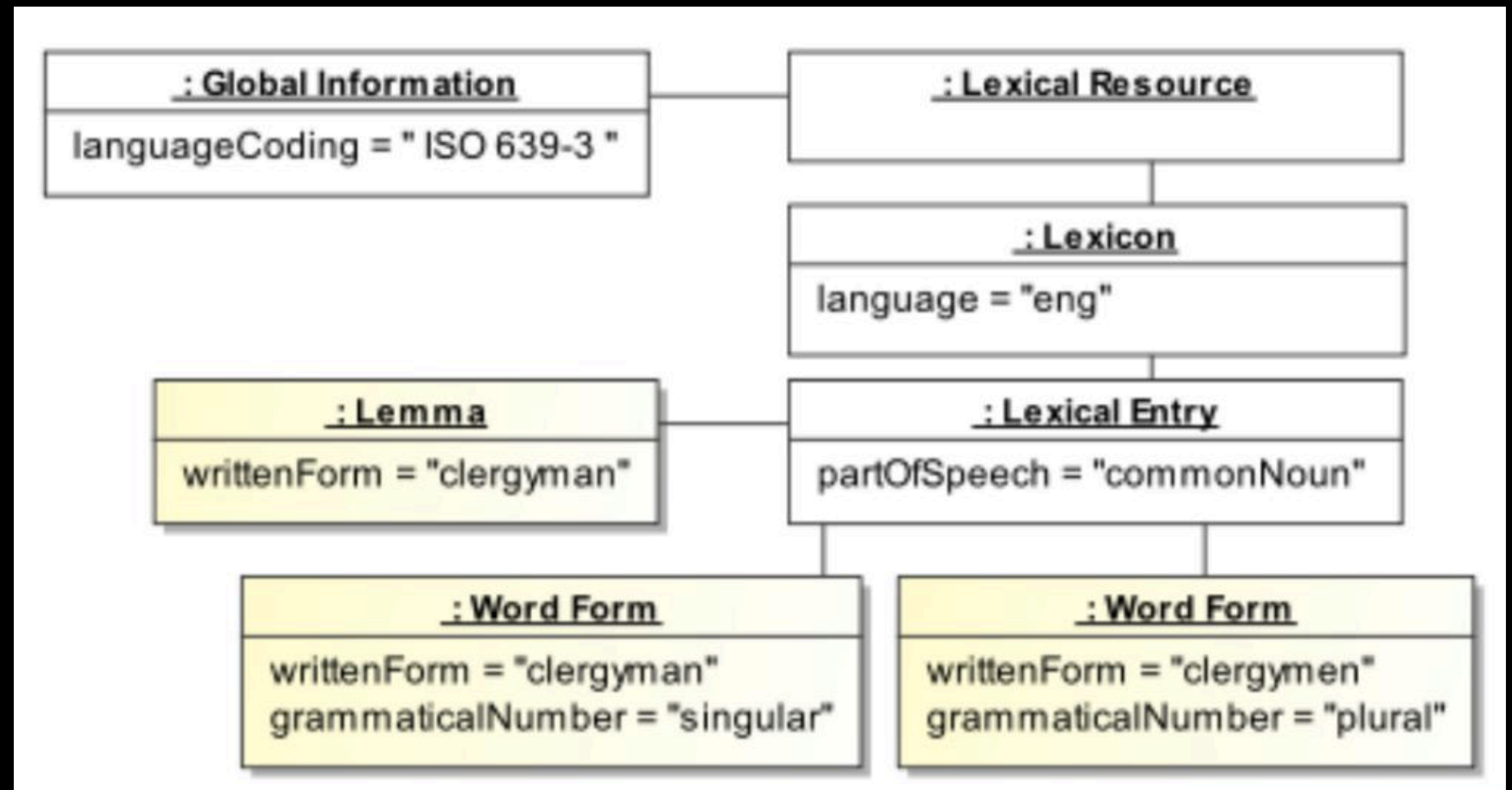
# Divadelní hry

```
<div type="Act" n="I">
  <head>ACT I</head>
  <div type="Scene" n="1">
    <head>SCENE I</head>
    <stage rend="italic"> Enter Barnardo and Francisco, two
    Sentinels, at several doors</stage>
    <sp>
      <speaker>Barn</speaker>
      <l part="Y">Who's there?</l>
    </sp>
    <sp>
      <speaker>Fran</speaker>
      <l>Nay, answer me. Stand and unfold yourself.</l>
    </sp>
```

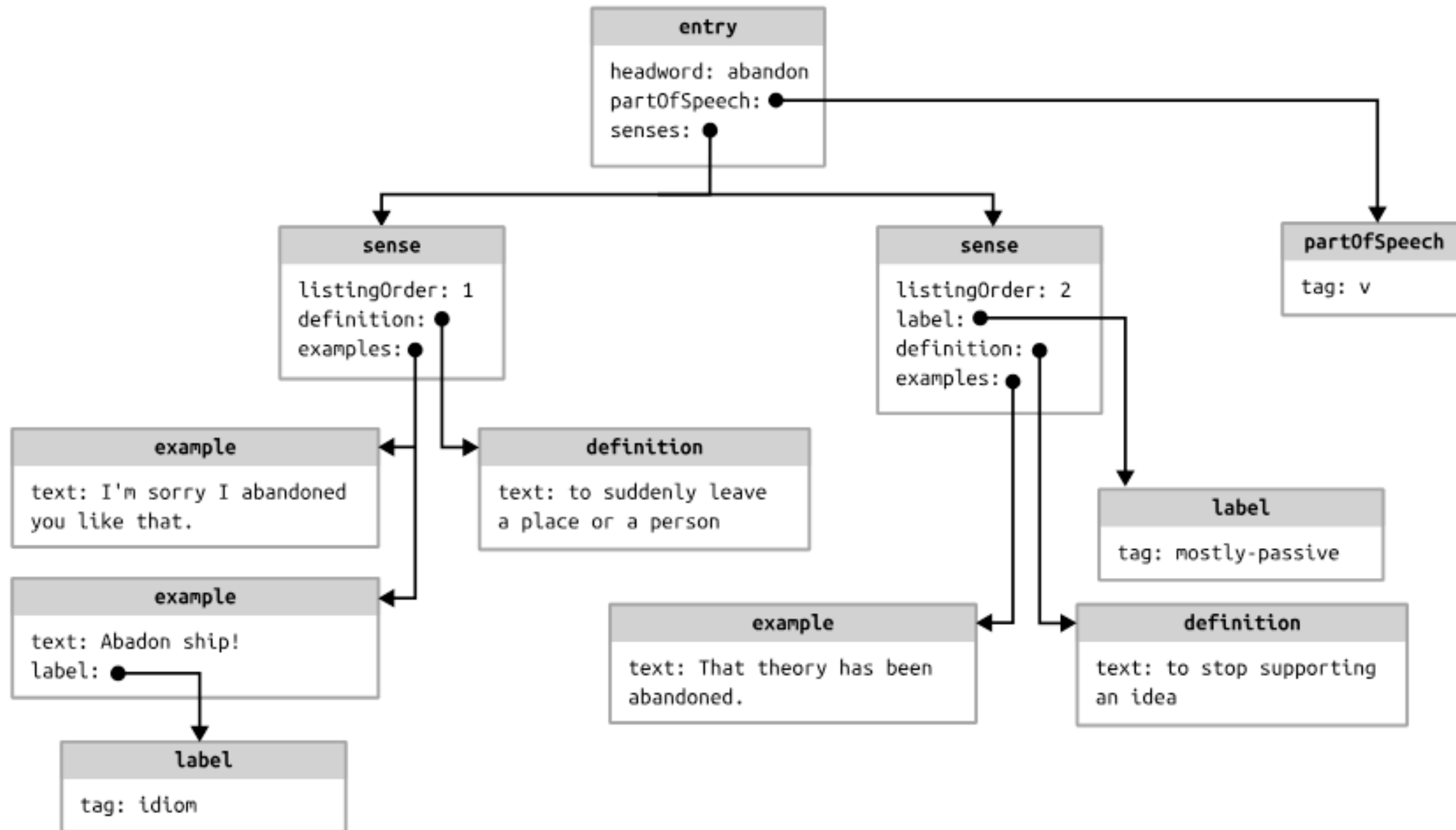


# Lexical Markup Framework

Jednotný model pro lexikální zdroje.



# DMLex



# NVH

headword: bank

partOfSpeech: noun

definition: an institution where you store or borrow money

translation: банка

example: I got a large loan from the bank.

translation: Я получил крупный кредит в банке.

definition: a stretch of land along a river

translation: берег

example: The house is on the north bank of the river.

translation: Дом находится на северном берегу реки.

## **Name-Value Hierarchy (NVH)**

Kratší, méně verbální.

Člověk: čtení i psaní.

Jednoduše převoditelný.

**Přístupy k datům: crowdsourcing**

# OPÁČKO: Crowdsourcing (širší)

1. Wiki model.

2. User-generated content.

3. Crowdsourcing (užší)

## Wiki model

- Specialisté na jednotlivá témata.
- Kontrola od jiných specialistů.
- Bez managementu. Seberegulace.



**User-generated content**

Od lidí pro lidi.

Malý management. Bez expertízy.

Náhodná, nekompletní data.

# User-generated content

Od lidí pro lidi.

The image displays three user-generated content cards arranged horizontally. Each card features a green checkmark icon and the word 'Expert' in a white box. The first card shows a hand holding a stylus over a tablet, with the text 'How to Take a Screenshot in Microsoft Windows'. The second card shows a man in a dark shirt and tie, with a green box indicating '12"', and the text 'How to Tie a Tie'. The third card shows a cartoon illustration of a man and a woman laughing, with the text 'How to Flirt'.

Expert

How to  
**Take a Screenshot in  
Microsoft Windows**

Expert

How to  
**Tie a Tie**

Expert

How to  
**Flirt**

## Crowdsourcing (užší)

Obrovská malá data od neexpertů...

... zpracovávaná experty.

Management.

# Crowdsourcing (užší)

Birdwatch,

COVID symptoms app,

English Dialect app,

...

# Přehled

Dictionary Writing Systems.

Markup + modeling.

Crowdsourcing.