Word at a Glance

a Customizable Word Profile Aggregator

Tomáš Machálek
Institute of the Czech National Corpus, Charles University
Motivation

● user-friendly presentation of word profiles based on available data
  ○ promotion of existing language resources
  ○ promotion of existing web applications that are linked from within Word at a Glance

● connecting different existing tools and datasets
  ○ within a single institution or project
  ○ across different sites

● intended audience
  ○ newcomers: easy access, showing the power of the data
  ○ advanced researches: overview of typical word's behaviour
The Application ((Word at a Glance) at a Glance)

- **Word/term knowledge aggregator**
  - meta-search approach,
  - focus on results interconnection,

- **Prepackaged application modules - tiles**
  - diverse data resources,
  - different presentation modes,
  - selection and combination of the tiles gives a WaG installation its character & look
The Application - Query Modes

1) Single word search

2) Two or more words comparison

3) Word translation
The Application - Tile

- basic building block
  - query result = set of tiles
- data resource + function + presentation

Already available: concordance, bar chart, pie chart, word forms, collocations, speeches, time-based distribution, word translations, word translations vs. contexts, geo areas, word freq. profile, filtered concordance, datamuse.com, raw HTML, multi-source freq. distribution
The Application - **Layout**

**Logical structure**

- Query mode M & primary language L
- Tile Group A
  - Tile A1
  - Tile A2
  - Tile A3
- Tile Group B
  - Tile B1
  - Tile B2

**Visual structure**

<table>
<thead>
<tr>
<th>Tile 1</th>
<th>Tile 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tile 3</td>
<td>Tile 4</td>
</tr>
<tr>
<td>Tile 5</td>
<td></td>
</tr>
<tr>
<td>Tile 6</td>
<td></td>
</tr>
<tr>
<td>Tile 7</td>
<td>Tile 8</td>
</tr>
</tbody>
</table>

*Mobile screen friendly*
The Application - Tiles Interaction

User query
1
Word database
2
Selected variant
3
Internal query
4

Tile 1
5
Tile 2
6
Tile 3

Tile 4

“At last the list was drawn up.”

“Far too much attention had been drawn to...”

“Patterson has drawn three conclusions about...”

“conclusions”, “list”, “attention”
Technical information

● built on existing development experience with KonText,
● written in TypeScript for both client and server side,
● based on React, RxJS, custom state management
● only system dependencies:
  ○ Node.JS + npm (package manager)
  ○ HTTP proxy server (Nginx, Apache,...) for WaG and attached services
● Configured via JSON files
  ○ JSON schema available for easier editing
Configuration (of a tile) - example

"CollocExamplesSynchronic": {
    "tileType": "ConcFilterTile",
    "label": {
        "cs-CZ": "Ukázky kolokací",
        "en-US": "Collocation examples"
    },
    "readSubqFrom": ["CollocationsSynchronic"],
    "apiURL": "http://kontext.korpus.test/kontext-api/quick_filter",
    "corpname": "syn2015",
    "posAttrs": ["word"],
    "metadataAttrs": [
        {
            "value": "doc.title",
            "label": {
                "cs-CZ": "Název",
                "en-US": "Title"
            }
        },
        {
            "value": "doc.author",
            "label": {
                "cs-CZ": "Autor",
                "en-US": "Author"
            }
        },
        {
            "value": "doc.biblio",
            "label": {
                "cs-CZ": "Bib. info",
                "en-US": "Bib. info"
            }
        }
    ]
}
Installation & deployment considerations

- **Backend services availability & reliability**
  - Multiple remote (3rd party) services => outages, processing times out of control
  - **solution**: caching, redundancy

- **Increased load on backend services**
  - Single WaG query produces typically at least N queries to backend services (N = num. of tiles)
  - **solution**: caching, load-balancing, proper backend service setup
Outlook

● Multi-word expressions
  ○ phrasemes, idioms, named entities,...

● Words comparison
  ○ = rewrite of an existing CNC application SyD

● More data resources
  ○ ElasticSearch, general SQL, NoSkE, Corpus Workbench,...

● 3rd party tiles / resource adapters
  ○ Add a custom module to the configuration -> compile -> deploy
Conclusion

● universal word meta-search tool,
  ○ usage not limited to linguistics,
● great for extracting interesting functions out of different tools and putting them together,
  ○ promoting of tools and data sets,
● production ready - https://www.korpus.cz/slovo-v-kostce/,
● open-source
  ○ developed on GitHub - https://github.com/czcorpus/wdglance,
● active development - more functions to come.
Thank you!