Big Data, Big Systems…

Roman Piontek
Director of SaaS Innovation

rpiontek@ebsco.com

Semantic Web

Where Art Thou?

Tim Berners-Lee: The „Semantic Web"
How to find structured and reliable information?

Finding information in libraries:

- Researchers spend a great deal of their time searching for content.

- **45% of digital library users express frustration in finding content.**

- 84% of ALL users start their search on Wikipedia or Google Scholar.

Google Scholar

☰ **Google** Scholar

educational media 🔍

◆ Articles

About 4.040.000 results (**0,09** sec)

Any time
Since 2019
Since 2018
Since 2015
Custom range...

Sort by relevance
Sort by date

☑ include patents
☑ include citations

✉ Create alert

[PDF] Analysis of **educational media** server workloads
JM Almeida, J Krueger, DL Eager… - Proceedings of the 11th …, 2001 - land.ufrj.br
This paper presents an extensive analysis of the client workloads for **educational media**
servers at two major US universities. The goals of the analysis include providing data for
generating synthetic workloads, gaining insight into the design of streaming content …
☆ 🙶 Cited by 273   Related articles   All 11 versions   »

Administering **educational media**: instructional technology and library services
JW Brown - 1972 - ERIC
Intended primarily as a college text for **media** specialists, librarians, audiovisual specialists,
and others who will assume positions of leadership in the field of **educational media**, this
book begins with an analysis of the place of the **media** program in the total **educational** …
☆ 🙶 Cited by 106   Related articles   All 2 versions   »

THE NEW **MEDIA**--MEMO TO **EDUCATIONAL** PLANNERS.
W Schramm - 1967 - ERIC
THIS STUDY REVIEWS THE **EDUCATIONAL** USES, EFFECTIVENESS AND COSTS OF
THE NEW **MEDIA** (RADIO, TELEVISION, PROGRAMED LEARNING, CORRESPONDENCE
STUDY, AND FILMS AND AUDIO-VISUAL AIDS) AND MAKES SUGGESTIONS TOWARDS …
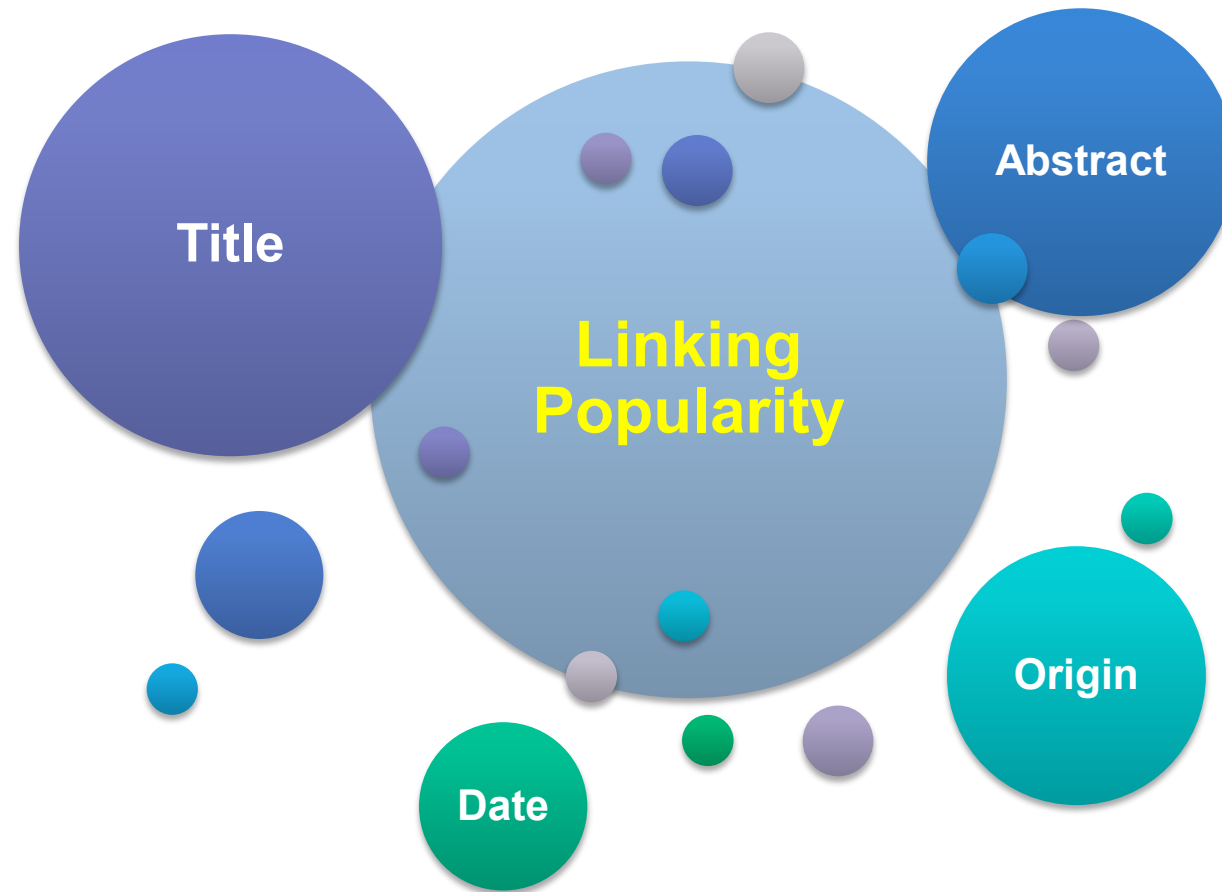☆ 🙶 Cited by 104   Related articles   »

[BOOK] The Economics of **Educational Media**.
L Wagner - 1982 - ERIC
This guide to the economic analysis of **media** use in education reviews the evidence on the
costs and benefits of **educational media** and provides a methodology for future decision-
making. It includes chapters on establishing a framework for **media** choice, evaluating the …
☆ 🙶 Cited by 66   Related articles   All 2 versions   »

What about subject terms?

Or subject indices?

Index Islamicus

CAB Abstracts

IET Inspec

HEINONLINE

PsycINFO
Available via EBSCOhost

THE Philosopher's INDEX

ATLA

rilm   Abstracts of
Music Literature

FSTA®
the food science resource

MathSciNet®

But even within the subject indices – subject headings can still differ substantially...

- *educational media* (ERIC)
- *educational resources* (GeoRef)
- *instructional materials* (MLA International Bibliography)
- *instructional media* (PsycINFO)
- *teaching aids & devices* (Education Abstracts, Education Source)
- *teaching materials* (MeSH/MEDLINE/PubMed, CINAHL)

Multilinguality!

EDS Thesaurus Visual Navigation ()
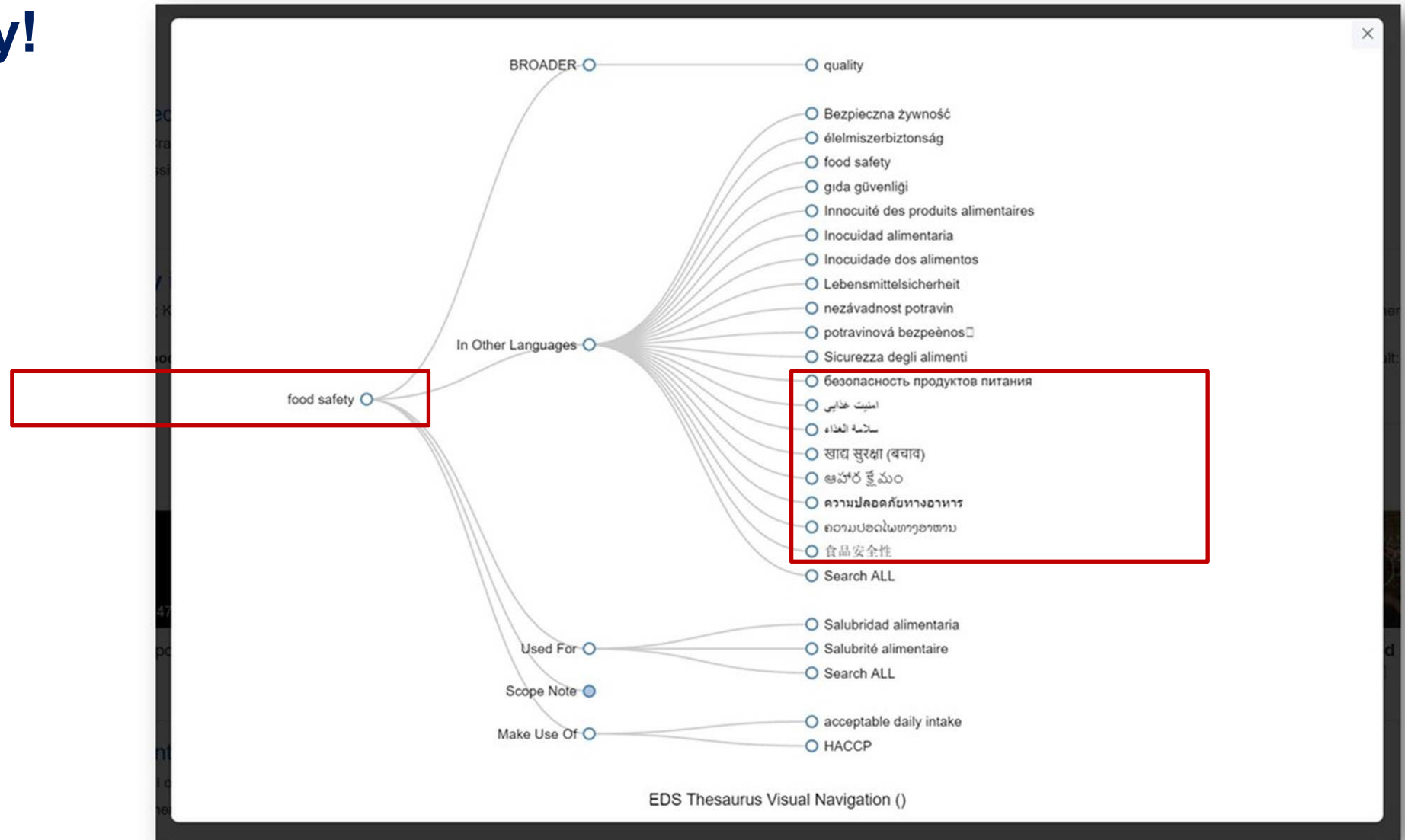
Thesauri, Taxonomies, MeSH...

Dostojewskij, Fedor Michailowitsch
Dostojewskij, Fedor Michajlovič
Dostojewskij, Fedor Michajlowitsch
Dostojewskij, Feodor
Dostojevskij, Fjodor Michajlovič
Dostoyeffsky, Fedor
Dostoyevsky, F.
Dostoyevsky, Feodor
Dostojewsky, Fedor M.
Dostojewsky, F. M.
Ntostogiephski, Phiontor Michaïlobits
Dāstavaskī, F. M.
Tosthojēwskhi, Th.
Dāstāyiwskī
Dosutoefusukī, F. M.
Dostojewski, Fjodor M. (RAK-ÖB)
Dostoievsky, F.-M.
Достоевский, Фёдор Михайлович

+ National Authority Files!

EBSCO

LINKED DATA

Ontologies: How to teach the WEB semantics?

- **Triples = Subject + Predicate + Object**
- **Proper description of concepts AND connections.**

Entity 1	Relationship	Entity 2
Engine	*is_Part_of*	Car
John Doe	*is_CoAuthor_with*	Sam Smith
Amsterdam	*is_A*	City

- **"RDF" is the descriptive IT standard for this.**
- **RDF corresponds to RDA in libraries.**

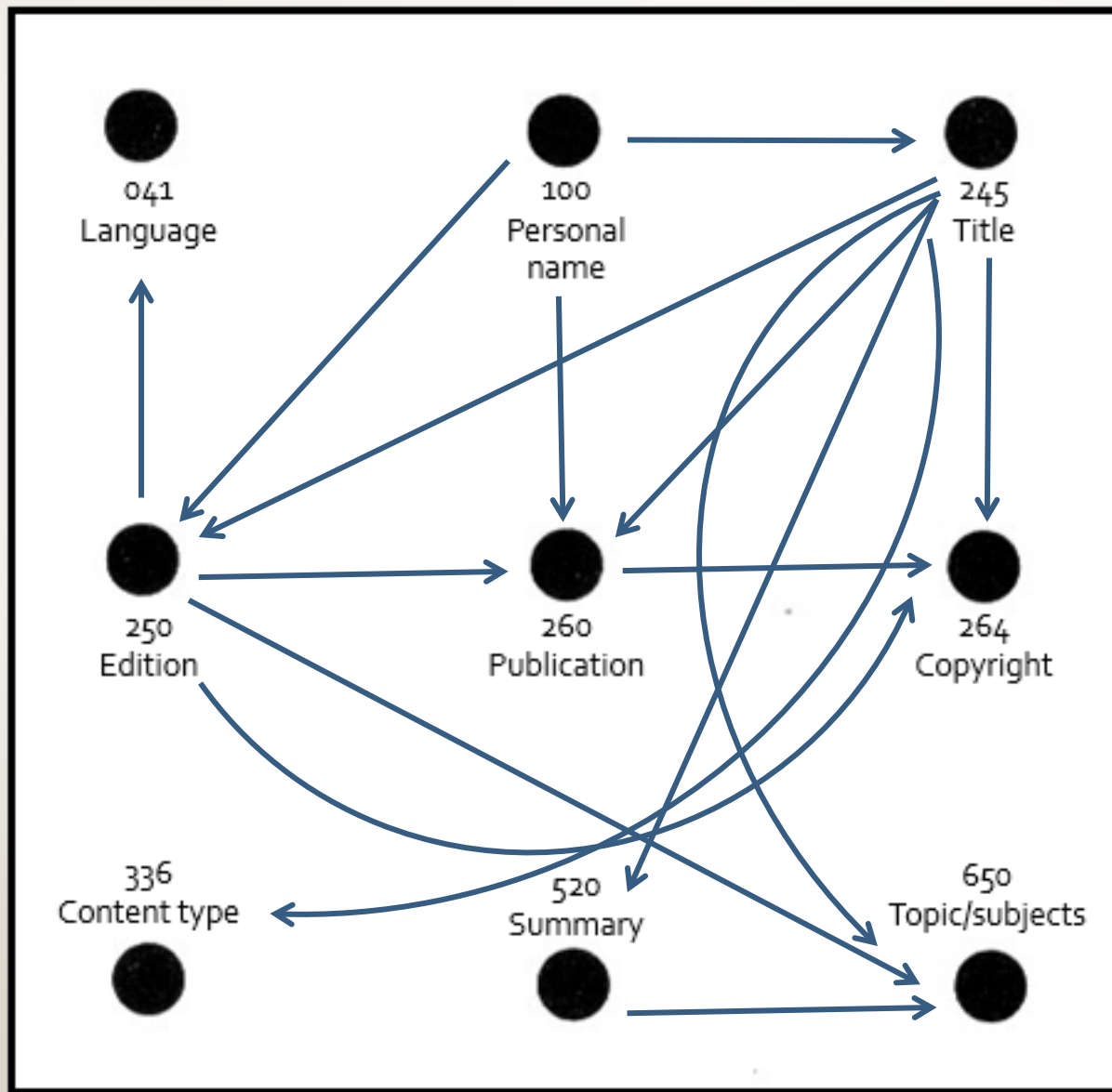**Many triples paint a picture**

"Triples" in the Library World:

Use of RDA FRBR Example

* "Imagine that you have a patron who needs a copy of Heaney's translation of Beowulf. She doesn't care who published it or when, only that it's Heaney's translation. What if you (or your patron) could place an interlibrary loan call on that expression, instead of looking through multiple bibliographic records... for multiple manifestations and then judging which record is the best bet on which to place a request? Combine that with functionality that lets you specify 'not Braille, not large print,' and it could save you time. Now imagine a patron in want of a copy, any copy, in English, of Romeo and Juliet." (Gonzalez, Linda. "What Is FRBR?" Library Journal 130 [Spring 2005 NetConnect]: 12 – 14).

* Example (using an RDF triple- you won't get what an RDF triple is yet but in later weeks you will –for now just check out the example)

expression	**Heaney's**	**translation of**	**Beowulf**
	Attribute- Group 2	relationship	Work patron wants

* Using this example, the work (i.e. the copy your library has) can be found by:

  * Translations of Beowulf
  * Heaney translations
  * Heaney AND Beowulf

  * The expression itself, Heaney's translation of Beowulf
  * Adding the attribute English translation can also be used

Each relationship is an additional access point.

Example of a "Web of Predicates":

- The 4th edition is in Spanish
- Maria Luguio wrote the book
- Maria Luguio wrote the books current edition
- Mario Luguio worked with the publisher XIX
- The title is in its 4th edition
- The title was published by XIX Inc.
- The title is under this copyright
- The title is a textual content type
- The title is about the summery (abstract)
- The 4th edition was produced in Italy
- XIX Inc. filed and received copyright
- The 4th edition was produced and in copyright by 1980
- The book is about horses
- The edition is about horse healthcare
- The summary describes horses

Visualisation

- According to a study conducted by the MIT Artificial Intelligence Lab, knowledge seekers preferred to use **browsing capabilities** over keyword search **40% of the time**.

artificial intelligence

Domain: Computing

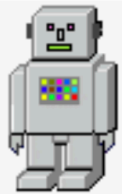**Add to search**

Group by
Relationships ▾

The branch of computer science that deal with writing computer programs that can solve problems creatively

artificial intelligence: has kind
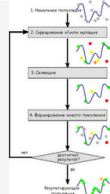
Computational intelligence

The expression computational intelligence usually refers to the ability of a computer to learn a specific ...

Explore concept   **Add to search**

Evolutionary computation

In computer science, evolutionary computation is a family of algorithms for global optimization inspired ...

Explore concept   **Add to search**

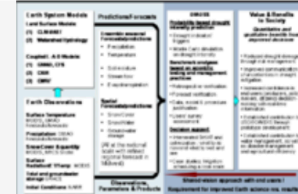### natural language processing

The branch of information science that deals with natural language information

Explore concept   **Add to search**

decision support system

A decision support system is an information system that supports business or organizational decision-...

Explore concept   **Add to search**

artificial intelligence: semantically related

Decision analysis

Theory of multiple intelligences

Chatbot

multi-agent system

ARTIFICIAL INTELLIGENCE
A program that can sense, reason,
act, and adapt

MACHINE LEARNING
Algorithms whose performance improve
as they are exposed to more data over time

**DEEP
LEARNING**
Subset of machine learning in
which multilayered neural
networks learn from
vast amounts of data

Limitations: e.g. Google's "Deep Dream"

What about **language**?

After all, most scientific data are couched in language.

Let's have a look at "big words" in English: (not that big for French or Italian speakers…)

- su**spic**ious
- incon**spic**uous
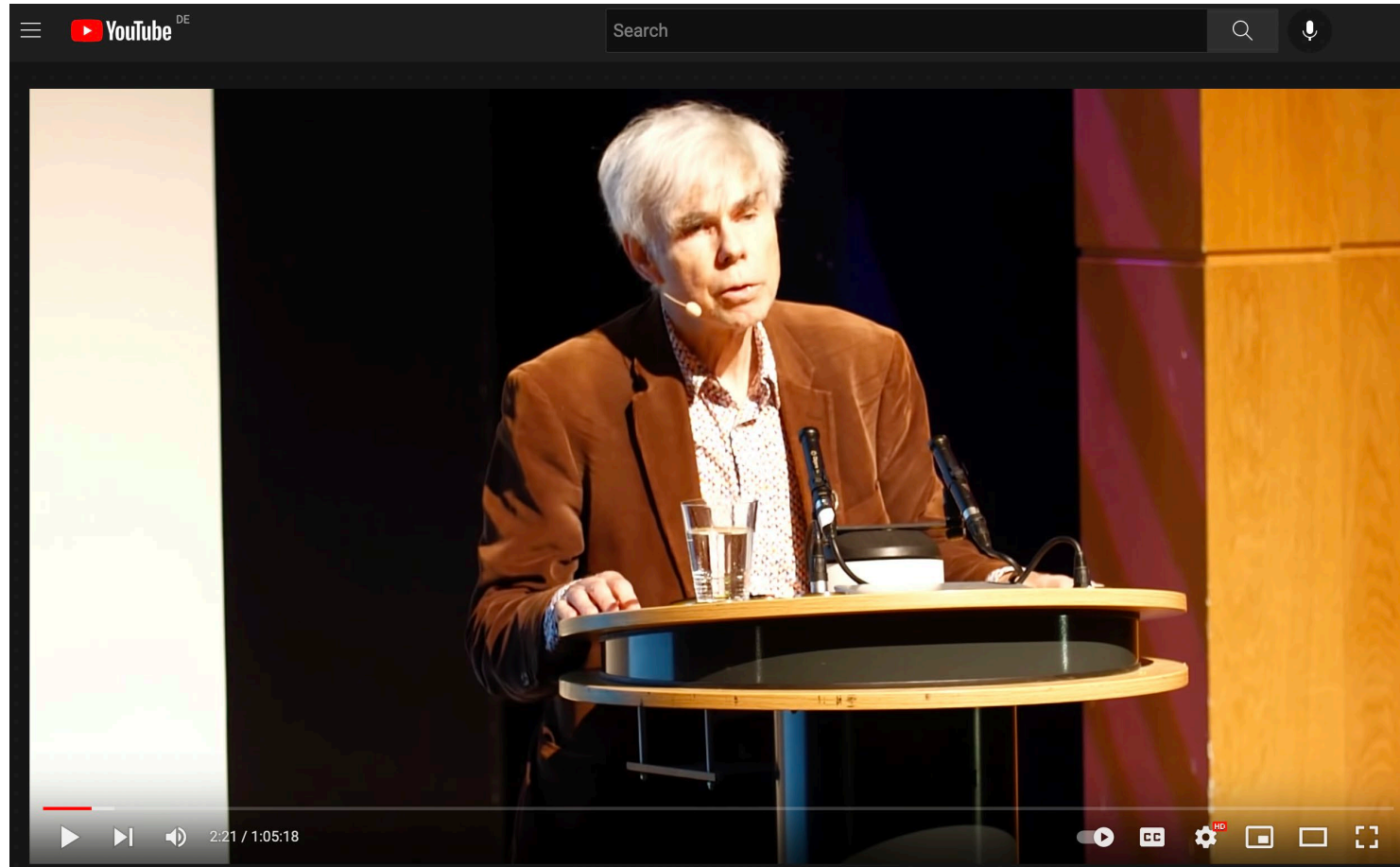- per**spic**acious
- au**spic**es

- **spec**tator
- **spec**tacular
- **spec**iation ← **spec**ies

All from the Latin root *spec- ~ spic-* = "to watch"

EBSCO

Would a machine be able to detect such patterns?

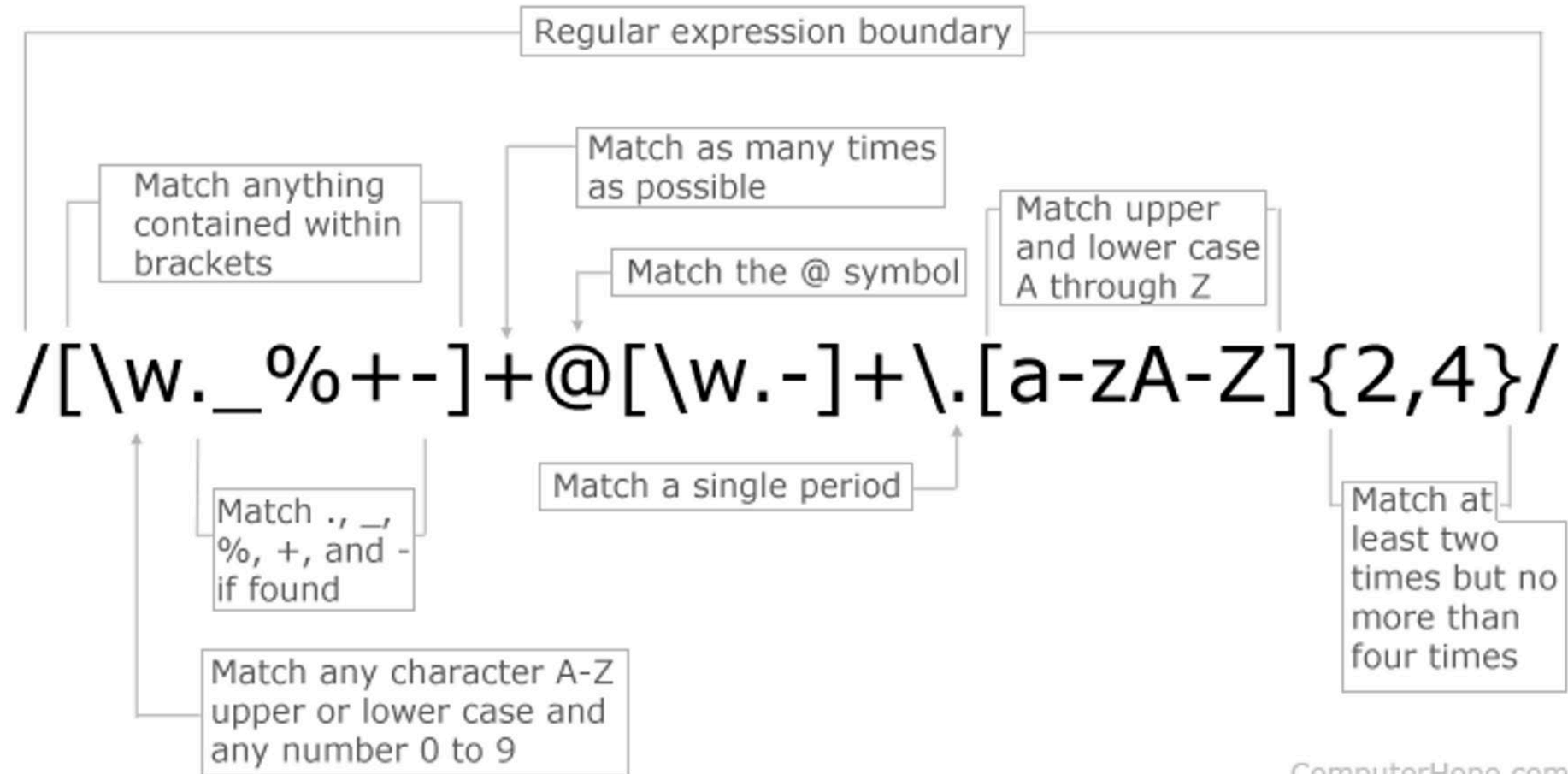**What about more complicated patterns in e.g. research data?**

Douglas R. Hofstadter:
"Reflections on Machine Translation"

Multiple plea:

- More **"data scientist"** in all of us, especially in libraries.

- BUT also awareness of the **limitations** of technology.

- I really recommend learning **"regular expressions"**.

Regular expressions are powerful tools for search, pattern recognition and rearrangement.

Regular expression boundary

Match anything contained within brackets

Match as many times as possible

Match the @ symbol

Match upper and lower case A through Z

/[\w._%+-]+@[\w.-]+\.[a-zA-Z]{2,4}/

Match ., _, %, +, and - if found

Match a single period

Match at least two times but no more than four times

Match any character A-Z upper or lower case and any number 0 to 9

ComputerHope.com

EBSCO

In conclusion:

Structure is key.

Sounds trivial—but it isn't.

Feel free to reach out to me if you would like to discuss:

- **Ontologies and Knowledge Graphs**

- **Language Processing, Internationalisation**

- **Search and Indexing**

- **Integrated Library Systems**

- **Open Source, Open Science, Open Access**

Thank you!

Roman Piontek
Director of SaaS Innovation

rpiontek@ebsco.com