

Úvod do organizace znalostí - řízené slovníky, ontologie

Josef Schwarz

22. 6. 2023



Spolufinancováno
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY





Úvod do organizace znalostí

Řízené slovníky, ontologie

PREZENTACE PRO ÚČASTNÍKY SÉRIE WEBINÁŘŮ EOSC-CZ

JOSEF SCHWARZ, CIT FF MU, LINDAT/CLARIAH-CZ

22. 6. 2023



„Jako ani ta nejbohatší knihovna, pokud není uspořádaná, neposkytuje tolik užitku jako velmi umírněná, ale dobře uspořádaná, právě tak i to největší množství poznatků, není-li propracováno vlastním myšlením, má mnohem méně ceny než daleko menší množství, avšak mnohostranně promyšlené.“

– ARTHUR SCHOPENHAUER, *PARERGA A PARALIPOMENA*, § 257
(1851; PŘEL. MILAN VÁŇA, 2011)

Východiska

Přesun od formálních otázek (struktura, metadata, formáty atd.) k problémům obsahovým (čeho se sledovaný informační objekt týká, o čem vypovídá, jaký je jeho obsah).

V této souvislosti je třeba zvažovat mj.:

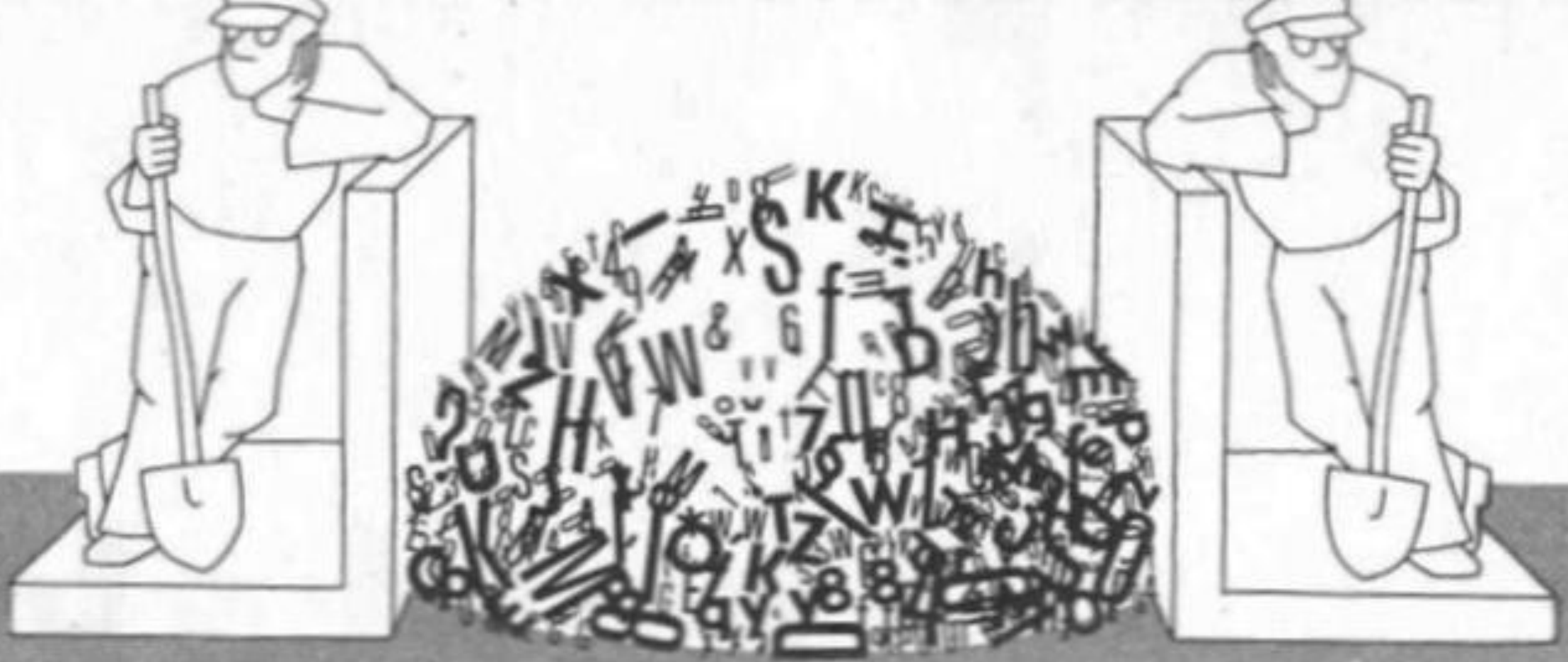
- Co je pojem?
- Co znamená třídít?
- Co vlastně třídíme a proč to děláme?
- Jak to děláme?
- Jak v tomto směru funguje lidská mysl a jak „mysl“ strojová?

Proč organizace znalostí?

NEBOLI: K ČEMU JE OBLAST V ANGLIČTINĚ NAZÝVANÁ
KNOWLEDGE ORGANIZATION DOBRÁ.



Photo by Valdemaras D. on Unsplash

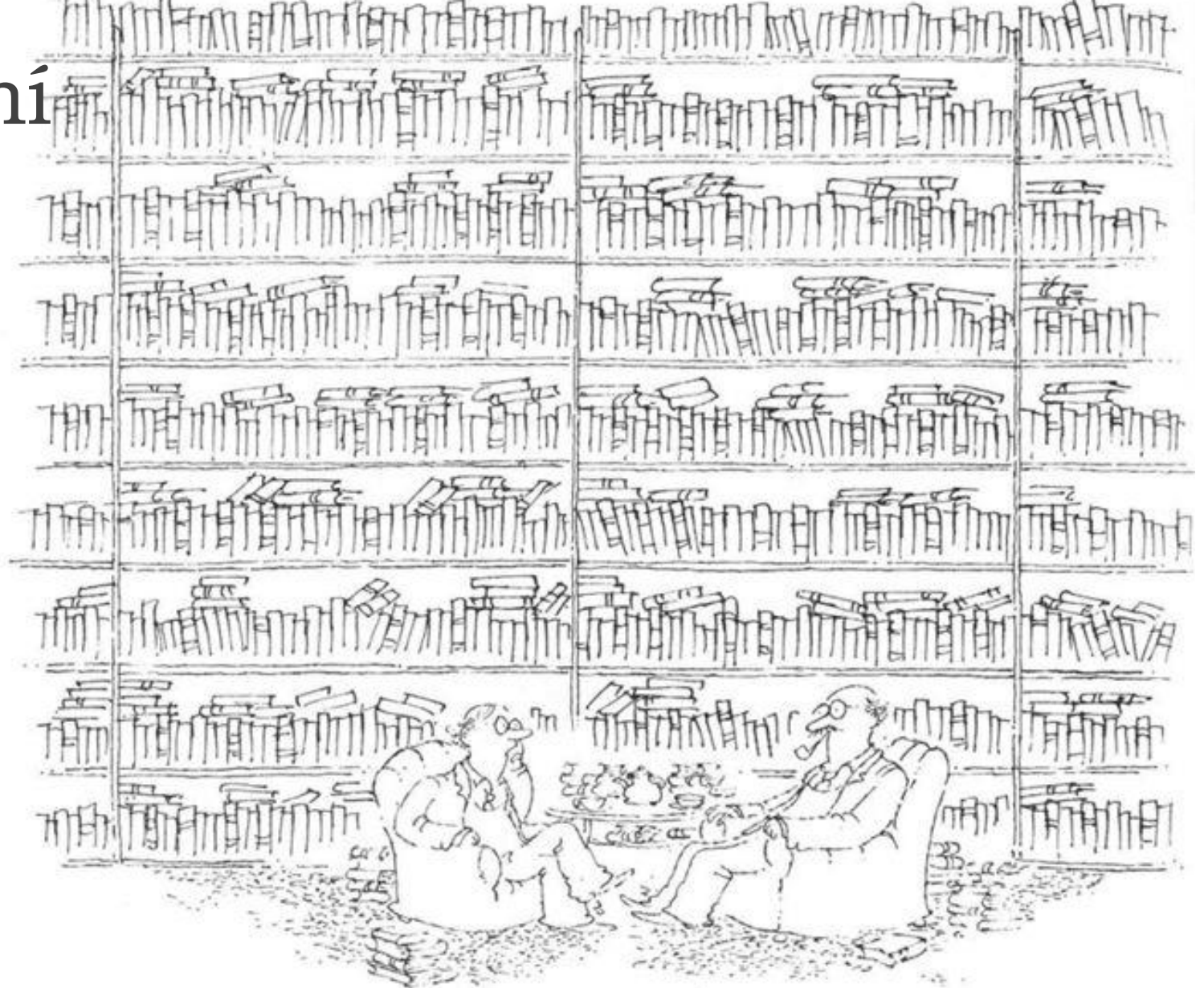


Informační exploze

Zvláště po druhé světové válce roste množství zachycených dat, informací, poznatků.

Informační přetížení

Objektivní důsledek: Informační exploze komplikuje efektivní vyhledávání informací i jejich využívání.

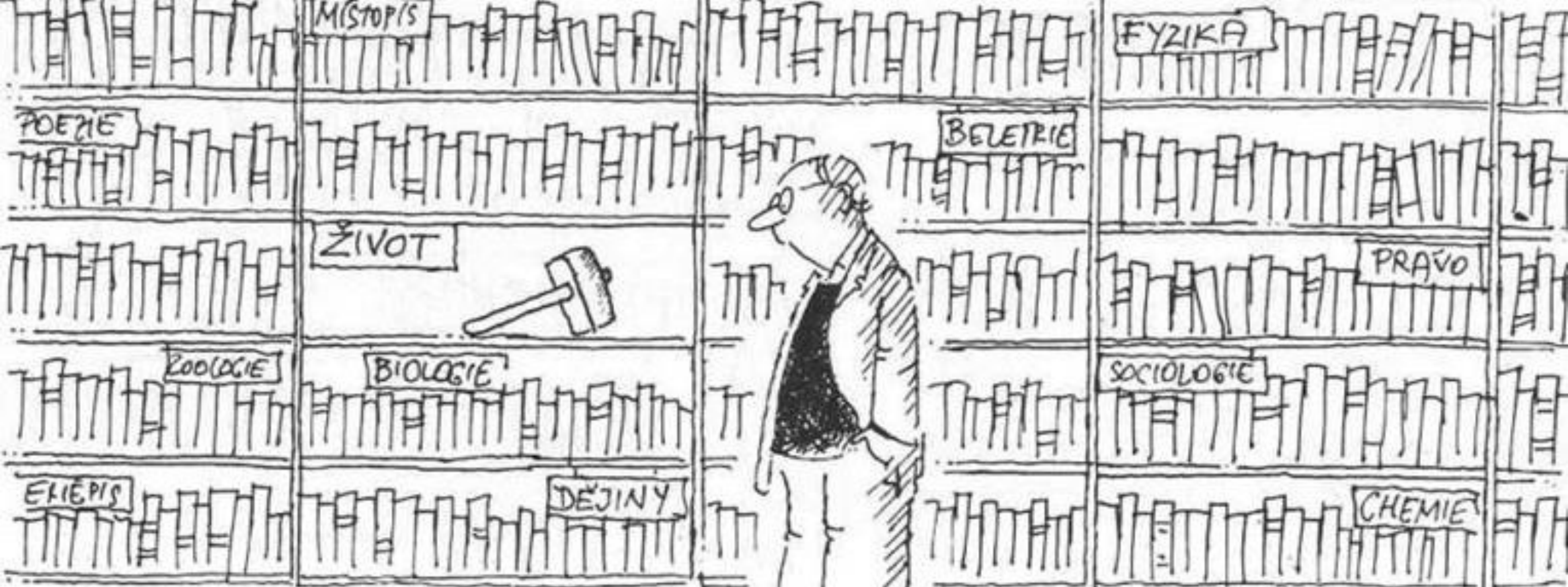


BOŽENA ... BOŽENA NĚMCOVÁ, ŘÍKÁTE ? HROME, TO JE MI NĚJAK POVĚDOMĚ.



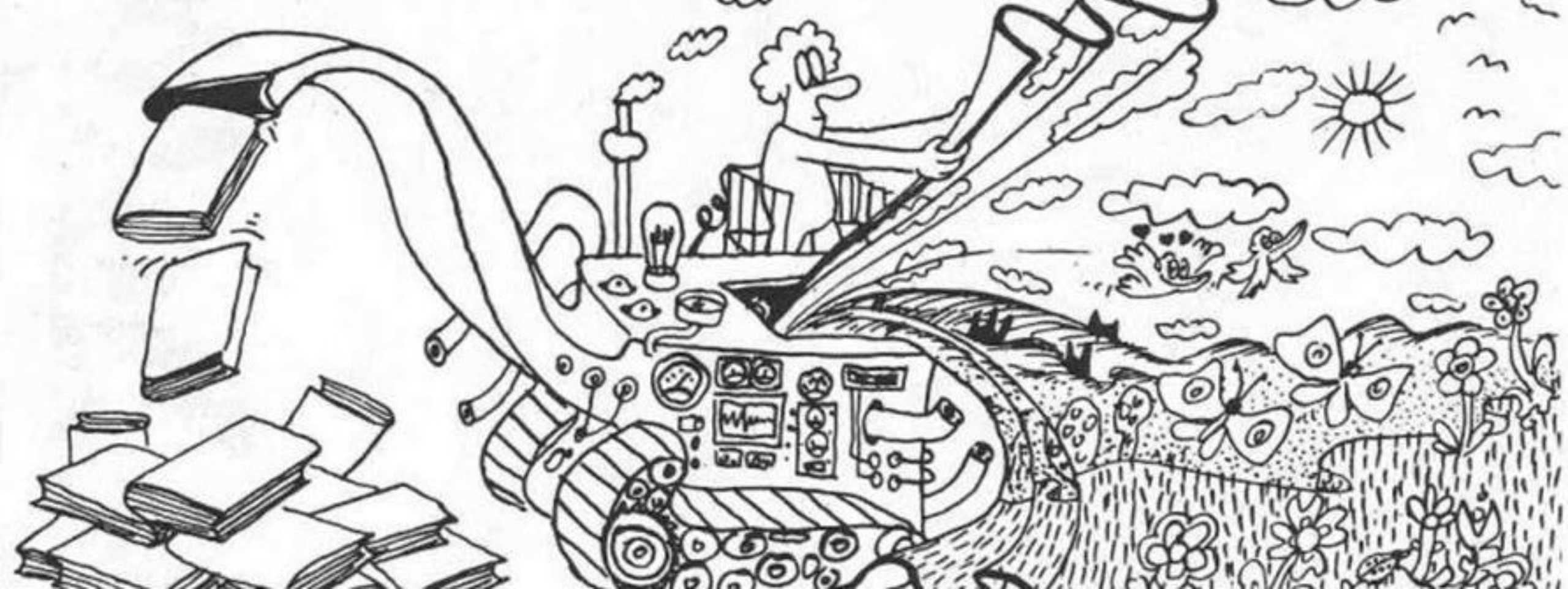
Informační neuróza/deprivace

Subjektivní důsledek: Psychologické dopady v podobě stavů způsobených nadbytkem informací nebo nedostatkem informací relevantních; vede k neschopnosti rychlého a racionálního rozhodování, ke stresu, únavě, roztržitosti a dalším negativním jevům.



Organizace znalostí

Jedním z řešení uvedených problémů je pořádání informací/dokumentů.



Specializace vs transdisciplinarita

Jsme skutečně schopni zachytit neustále sílící proud zaznamenaného poznání? A jaká je vlastně povaha tohoto poznání? Jsou systémy organizace znalostí schopny se tomuto rychlému tempu přizpůsobovat?

Vnitřní pořádací princip KOS

předurčuje možnosti a limity inference

ABECEDNÍ ŘAZENÍ

číselníky
folksonomie
grafy znalostí
hesláře
jmenný prostor
klasifikace
klastry
klíčová slova
konspekt
mapy témat
ontologie
pojmové grafy
předmětová hesla
rejstříky
rotované rejstříky
rubrikátory
řízené slovníky
selekční jazyky
sémantické sítě
seznamy
soubory autorit
štítky
taxonomie
tezaury
unitermy

??? ÚKOL

štítky
hesláře
klastry
seznamy
tezaury
konspekt
unitermy
číselníky
ontologie
rejstříky
taxonomie
mapy témat
folksonomie
klasifikace
rubrikátory
klíčová slova
pojmové grafy
grafy znalostí
jmenný prostor
řízené slovníky
selekční jazyky
sémantické sítě
soubory autorit
předmětová hesla
rotované rejstříky

TEMATICKÉ ŘAZENÍ

01 GENERICKÉ POJMY

jmenný prostor
ontologie
řízené slovníky
selekční jazyky
sémantické sítě

05 ŘÍZENÉ SLOVNÍKY

05.1 ŘÍZENÉ SLOVNÍKY PŘEDMĚTOVÉHO TYPU

folksonomie
hesláře
klíčová slova
předmětová hesla
rejstříky
rotované rejstříky
seznamy
soubory autorit
štítky

tezaury
unitermy

05.5 ŘÍZENÉ SLOVNÍKY SYSTEMATICKÉHO TYPU

číselníky
klasifikace
klastry
konspekt
rubrikátory
taxonomie

10 FORMÁLNÍ ONTOLOGIE

grafy znalostí
mapy témat
pojmové grafy

TROCHA TEORIE I.

Několik základních pojmů

V JAKÉM VZTAHU JSOU DATA, INFORMACE A ZNALOSTI? JAK SE LIŠÍ DATA, METADATA A PARADATA? A PROČ V KNIHÁCH NENAJDEME ZNALOSTI...

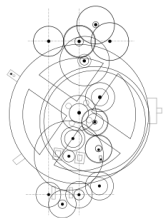
Data, informace, znalosti

Data

- Vymezení: Zaznamenané symbolické reprezentace reálného světa nebo světů fikčních
- Médium (nosič): Analogové (hlína, kámen, papír atd.) nebo digitální
- Struktura: Data strukturovaná (seznamy, tabulky, databáze atd.) nebo nestrukturovaná (text)
- Formát: Číselný, textový, obrazový, zvukový, audiovizuální, kombinovaný

Příklady:

- [460.8](#)
- [NA1995](#)
- [Příštpo](#)



Informace

- Vymezení: data s přiděleným významem
- Médium (nosič): Analogové (hlína, kámen, papír atd.) nebo digitální
- Struktura: Informace faktografické, plnotextové, bibliografické atd.
- Formát: Číselný, textový, obrazový, zvukový, audiovizuální, kombinovaný

Příklady:

- 460,8 m.n.m.
- NA1995 Architecture as a profession
- obec Příštpo

Znalost

- Vymezení: informace s přiděleným kontextem
- Médium (nosič): Mozek
- Struktura: Poznatkový tezaurus
- Formát: Elektrochemický signál
- Druhy: explicitní, tacitní

Příklady:

- 460,8 m.n.m. (Říp) **ÚKOL 1**
- NA1995 Architecture as a profession (LCC)
- Příštpo u Jaroměřic nad Rokytnou

Exkurz: data, metadata, paradata

Data

Metadata

- Vymezení: Data o datech související s jejich obsahem, vlastnostmi, formátem a dalšími charakteristikami.
- Dělení: popisná, strukturální, technická, administrativní aj. (záleží na kritériu členění)
- Příklady:
 - 460,8 m.n.m. | [Bpv](#)
 - NA1995 Architecture as a profession | Last revision 2013/30/12
 - Příklad u Jaroměřic nad Rokytnou | [544752](#)

Paradata

- Vymezení: Údaje o procesu vzniku/sběru, použití, interpretaci a kvalitě dat a jeho metodách; zajišťuje správné pochopení dat, jejich adekvátní využití a reprodukovatelnost výzkumu
 - 460,8 m.n.m. || Ověřeno vlastním měřením
 - NA1995 Architecture as a profession || Duplicitní k TU234B
 - Příklad u Jaroměřic nad Rokytnou || **Definitivní stav**

Data, informace, znalosti

Faktograficko-terminologická poznámka

POZOR!!!:

- Neexistují tzv. **tvrdá data** ve smyslu nezpochybnitelných údajů, protože data jsou výsledkem výzkumu, měření a interpretace za použití specifických metrik, nástrojů a metodologie. *Příklad 1 ([Mt. Everest](#))*
- Nezaměňovat **fakta a data** – fakta jsou údaje nebo tvrzení, která lze ověřit a jsou považována za pravdivá; data naproti tomu jsou pouhým záznamem bez pravdivostní hodnoty.

Jazyková poznámka:

Termín	Singulár	Plurál
fakt*	fakt, faktum	fakta
dat*	[neexistuje] (údaj)	data (údaje)
informac*	informace	informace
znalost*	znalost („jednotková“ znalost: poznatek)	znalosti

TROCHA TEORIE II.

Sémiotický trojúhelník

ZÁKLADNÍ EPISTEMOLOGICKÝ MODEL A JEHO SOUVISLOST S FILOZOFICKÝM MODELEM K. R. POPPERA. OBJEKTIVNÍ A SUBJEKTIVNÍ KLASIFIKACE. DŮSLEDKY PRO ORGANIZACI ZNALOSTÍ: CO VLASTNĚ TŘÍDÍME?

Sémiotický trojúhelník a teorie tří světů Karla Poppera a Johna Ecclese

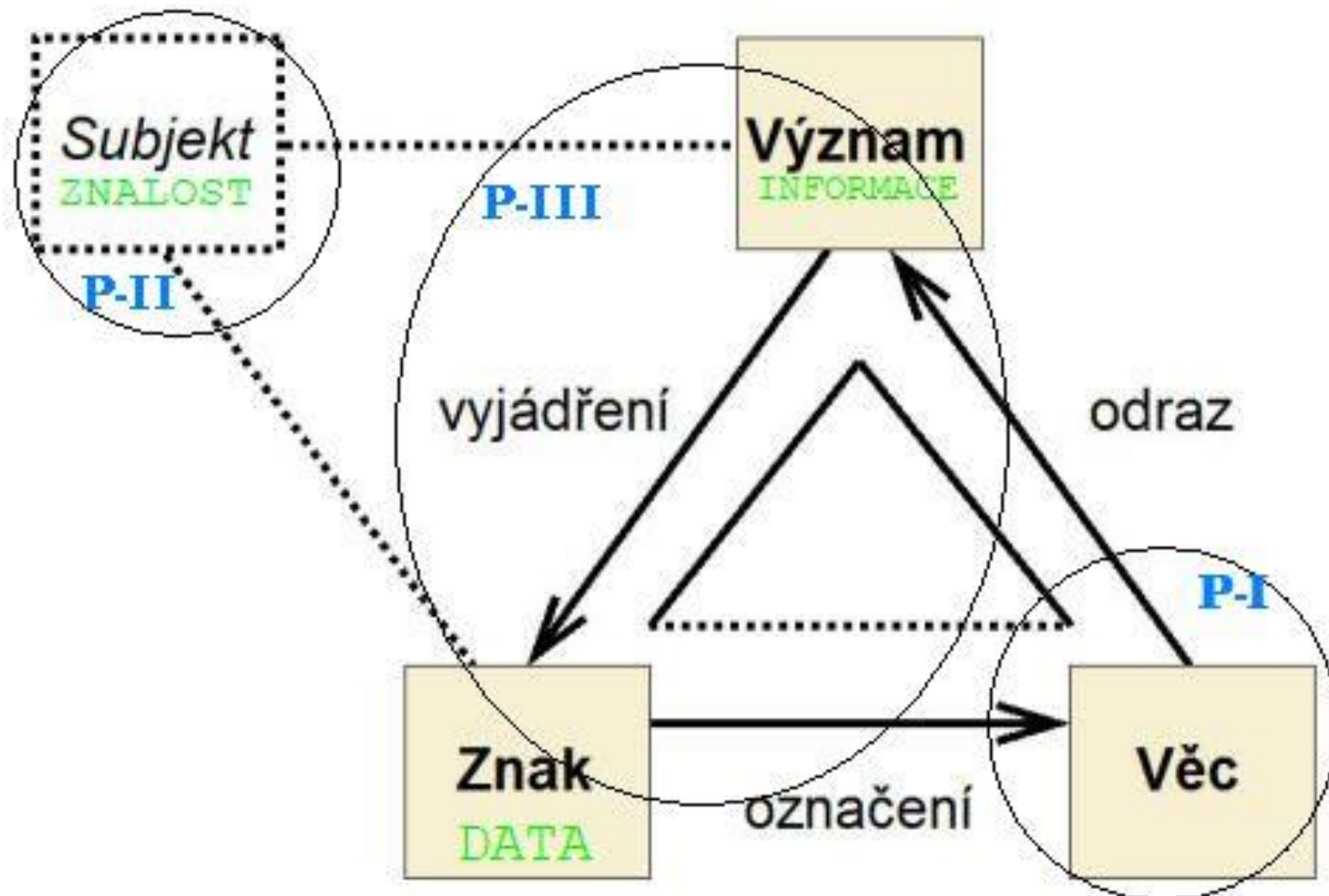


Schéma převzato z KUČEROVÁ, Helena. Co analyzujeme při obsahové analýze dokumentů? K pojmu aboutness v organizaci znalostí. Knihovna [online]. 2014, roč. 25, č. 1, s. 36-54 [cit. 2023-03-15]. Dostupný z WWW:

<http://knihovna.nkp.cz/knihovna141/141036.htm> a upraveno.

Popperův svět III

Typologie a atributy

Zaznamenané poznání

- Informační objekt
 - Dokument
 - Vědecká zpráva
 - Časopis
 - Vědecký časopis
 - Společenský časopis
 - Znalecký posudek
 - Faktura
 - 3D objekt
 - Kamenná stéla

Atributy informačního objektu

ÚKOL 2: Co je komiks?

Atributy informačního objektu

Označení	Dokumentografická úroveň	Příklad1	Příklad2	Příklad3	Příklad4	Příklad5	Příklad6
		Informační objekt: Trajánův sloup	Informační objekt: Enúma eliš	Informační objekt: Babička	Informační objekt: Zpívá Jan Werich	Informační objekt: Filmové Brno	Informační objekt: Když se země chvěje
A	nosič/médium	kámen	hlína	papír	analogové médium	digitální médium	digitální médium
B	forma	kamenný sloup	hliněné tabulky	kniha	gramofonová deska	databáze	Webová stránka
C	typ/druh	pamětní sloup	klínopisné tabulky	tištěná kniha	LP (dlouhohrající gramofonová deska)	relační databáze	Webový dokument
D	formát obsahu	text	text	text	zvuk	data	komiks
E	žánr	pamětní reliéfy	epos	beletrie /próza	písně	faktografie	naučná fikce
F	obsah	římsko-dácké války	kosmogonické mýty	venkovský život	život, kultura, společnost	brněnská kina	Zemětřesení, seismologie

Co třídí (organizuje):

BADATEL (ORNITOLOG)

A/ Pozorování a analýza/syntéza jevu: *datel břidlicový* (věc)

B/ Třídění jevu za pomoci taxonomie: *datel břidlicový* (význam/znak)

C/ Záznam o jevu: *datel břidlicový* (informační objekt)

INFORMAČNÍ SPECIALISTA

A/ Pozorování a analýza/syntéza jevu: *datel břidlicový* (informační objekt)

B/ Třídění jevu za pomoci systému organizace znalostí: *datel břidlicový* (význam/znak)

C/ Záznam o jevu: *datel břidlicový* (metadata)

DIGITÁLNÍ BADATEL

A/ Pozorování a analýza/syntéza jevu: *datel břidlicový* (metadata)

B/ Třídění jevu za pomoci systému organizace znalostí a dalších nástrojů pro zpracování (velkých) dat a metadat: *datel břidlicový* (metavýznam/metaznak)

C/ Záznam o jevu: *datel břidlicový* (metametadata)

// Pauza s úkoly

ÚKOL 3: Je těhotenství nemoc?

ÚKOL 4: Je Franz Kafka český spisovatel?

ÚKOL 5: Je nealkoholické pivo nealkoholický nápoj?

// Pauza s úkoly – autorské řešení

ÚKOL 3: Je těhotenství nemoc?

- [Ano: MKN, třída O](#)
- [Ne: PSH, třída bi \(biologie\)](#)

ÚKOL 4: Je Franz Kafka český spisovatel?

- Ano: narodil se v Praze, tvořil v českém prostředí
- Ne: psal německy (a není tudíž v [LČL](#))

ÚKOL 5: Je nealkoholické pivo nealkoholický nápoj?

- Ano: jmenuje se tak
- Ne: obsahuje zbytkový alkohol (u nás cca 0,5 %, [v jiných zemích odlišně](#))

Třídění jako objektivní proces?

„Prekurzory“ třídění:

- ❑ ontologická východiska (co je svět?)
- ❑ epistemologická východiska (realismus X nominalismus | konceptualismus)
- ❑ teleologický východiska (účel)

Teoretické přístupy k třídění:

- ❖ objektivistické stanovisko (bere v potaz jen Popperovy světy I a III):
 - ❖ klasická teorie kategorií
- ❖ experiencialistické stanovisko (zohlednění i Popperova světa II):
 - ❖ kognitivní vědy (prototypy a bázové kategorie)
 - ❖ doménová analýza

Základní lingvistické pojmy

- ❑ komunikace, lingvistika, lexikografie, lexikologie, terminologie, sémiotika
- ❑ jazyk (přirozený jazyk, umělý jazyk, formální jazyk), metajazyk
- ❑ symbol, znak, pojem (intenze, extenze, denotát), význam, smysl, kód, sémantický trojúhelník, arbitrárnost znaku
- ❑ gramatika, morfologie, syntax
- ❑ lexikum, lexikální jednotka, slovo, sousloví, inverze, termín
- ❑ homonymie, homofonie, homografie, polysémie, synonymie, antonymie
- ❑ hyponymie, hyperonymie, kohyponymie, meronymie, holonymie

Výklad těchto i dalších termínů v [databázi TDKIV](#).

Složky KOS

1. Slovník (řízený slovník, pořádací znaky) a jeho struktura (vztahy mezi pořádacími znaky)
 - Hierarchické vztahy (rododruhovové/generické; partitivní/celek-část)
 - Asociační vztahy (věcná souvislost)
 - Ekvivalenční vztahy (preferované a nepreferované jednotky)
2. Morfologie (pravidla pro tvorbu pořádacích znaků; např. inverze sousloví, binomická nomenklatura, základ a vlastnosti notace)
3. Syntax (pravidla pro spojování pořádacích znaků, tedy pravidla pro popis informačního objektu a pro formulaci rešeršního dotazu při vyhledávání)

Typy KOS I

Principium divisionis: Hlavní účel KOS

POPIS INFORMAČNÍCH OBJEKTŮ

selekční jazyky, řízené slovníky

- [MDT](#), [PSH](#), [LCSH](#), [AAT](#), [Agrovoc](#), [CZ-NACE](#)

REPREZENTACE POJMŮ

taxonomie, ontologie

[WordNet](#), [CRM-CIDOC](#), [OOPS](#), [DoCO](#)

Typy KOS II

Principium divisionis: Cílové atributy informačního objektu

IDENTIFIKAČNÍ (FORMÁLNÍ) ÚDAJE

[Soubor jmenných autorit](#)

**VĚCNÉ (OBSAHOVÉ, PŘEDMĚTOVÉ,
TEMATICKÉ) ÚDAJE**

[MDT](#), [PSH](#), [LCSH](#), [AAT](#), [Agrovoc](#), [CZ-NACE](#)

Typy KOS III

Principium divisionis: Charakter řízeného slovníku

ENUMERATIVNÍ CHARAKTER

[CZ-NACE](#)

FAZETOVÝ CHARAKTER

[CC \(Dvojtečkové třídění; Ranganathan\)](#)

[MDT, PSH, LCSH, AAT, Agrovoc](#)

Typy KOS IV

Principium divisionis: Uspořádání řízeného slovníku

ABECEDNÍ USPOŘÁDÁNÍ

[LCSH](#), [AAT](#), [Agrovoc](#), [PSH](#)

SYSTEMATICKÉ (LOGICKÉ) USPOŘÁDÁNÍ

[MDT](#), [IPC](#), [ICS](#)

Typy KOS V

Principium divisionis: Stupeň formalizace pořadacích znaků

POŘÁDACÍ ZNAKY ZALOŽENÉ NA
PŘIROZENÉM JAZYCE

[LCSH](#), [AAT](#), [Agrovoc](#), [PSH](#)

POŘÁDACÍ ZNAKY ZALOŽENÉ NA
FORMÁLNÍM KÓDU (NOTACI)

[MDT](#), [IPC](#), [ICS](#)

Systemy organizace znalostí II

Reprezentace KOS

JAK VZTAHY MEZI PRVKY KOS ZOBRAZUJEME A VIZUALIZUJEME?

Vizualizace pojmových vztahů

[Mapy témat](#) ČSN ISO/IEC 13250 (36 9830)

[Konceptuální grafy](#)

[Grafy znalostí 1](#)

[Grafy znalostí 2 \(CRM-CIDOC\)](#)

Systemy organizace znalostí III

Vstupní zpracování informačních objektů

CO ZNAMENÁ INDEXOVAT A KLASIFIKOVAT? V JAKÉ ÚROVNI
ZPRACOVÁNÍ SE POHYBUJEME?

Věcné zpracování informačního objektu

INDEXACE

Princip:

Popis objektu souborem pořadacích znaků, které jako celek vyjadřují obsah.

[Příklad](#)

KLASIFIKACE

Princip:

Zařazení objektu na určité místo v informačním systému.

[Příklad 1](#)

[Příklad 2 \(dále\)](#)

159.96-T



17-SO

- 159.96 Zvláštní duševní stavy a procesy. Hlubinná psychologie osobnosti
Special mental states and processes.
Depth psychology
- 159.964.2 Psychoanalýza
Psychoanalysis
- 159.97 Abnormální psychologie
Abnormal psychology
- 159.98 Aplikovaná psychologie
Applied psychology
- 16 Logika
Logic
- 165 Teorie poznání.
Epistemologie
Theory of knowledge. Epistemology
- 17 Morální filozofie. Etika
(Ethics (moral philosophy))

159.922-K



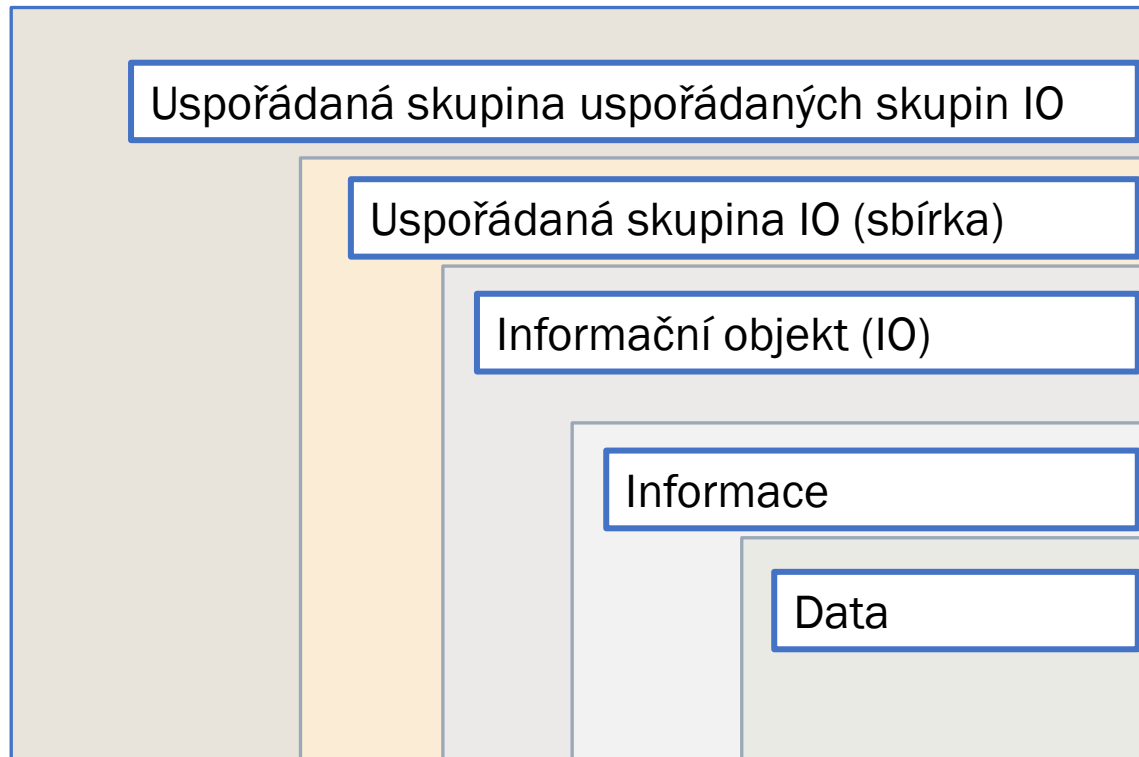
159.96-Ž

- 159.922 Vývojová psychologie, národní mentalita
Developmental psychology, national characteristics
- 159.923 Psychologie osobnosti
Psychology of personality
- 159.925 Analýza výrazu.
Grafologie
Graphology
- 159.93 Cítění.
Smyslové vnímání
Sensation. Sensory perception
- 159.94 Výkonové funkce.
Emoce
Executive functions. Emotions
- 159.95 Vyšší duševní procesy.
Hlubinná psychologie
Higher mental processes
- 159.96 Zvláštní duševní stavy a procesy. Hlubinná psychologie osobnosti
Special mental states and processes.
Depth psychology

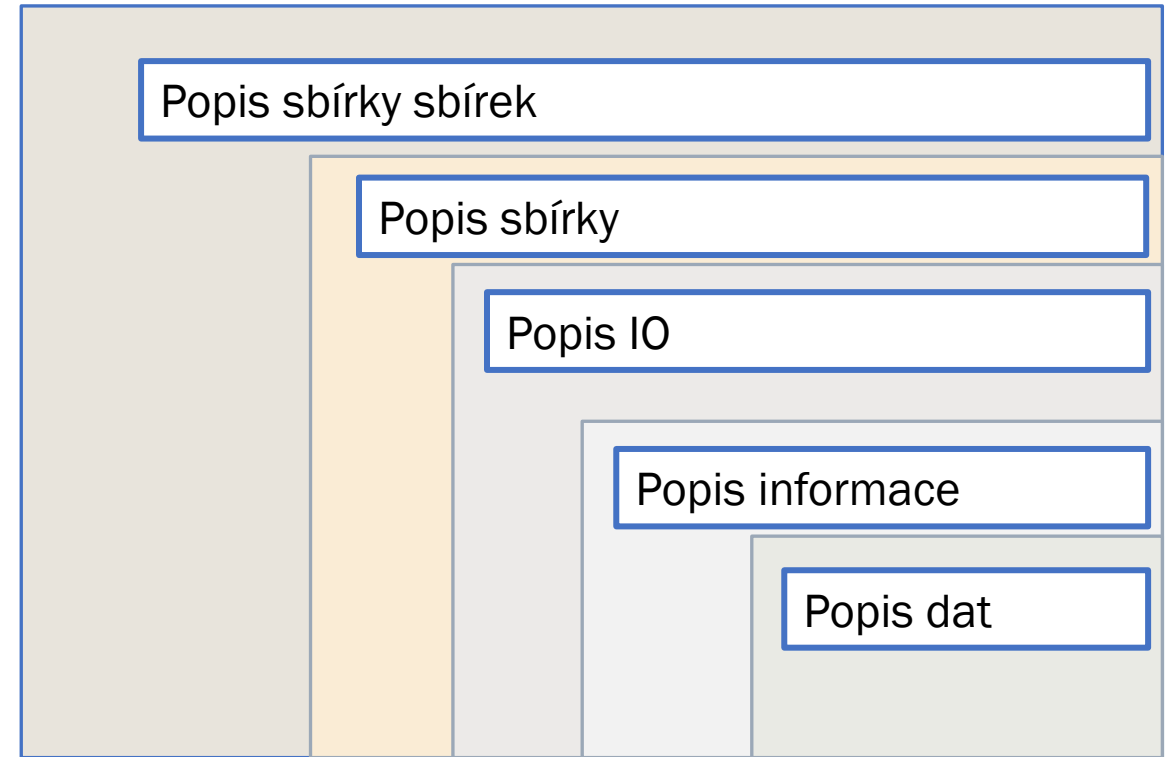


Úrovně zpracování inf. objektu

Primární úroveň (data, plný text)



Sekundární úroveň (metadata)



Systemy organizace znalostí IV

Normalizační funkce KOS

CO POTŘEBUJEME VÍCE?: JEDNOTU, NEBO EXPRESIVITU?

Funkční omezení normalizace

Normalizace (standardizace) vs reprezentace

HUSTOPEČE (U BRNA)

CISOB

PH NK

TGN

HUSTOPEČE NAD BEČVOU

CISOB

PH NK

[TGN 0]

Hypotéza: V TGN nejsou Hustopeče nad Bečvou s největší pravděpodobností proto, že nejsou město, ale pouhý městys.

Pro zájemce o další studium

- ❖ KUČEROVÁ, Helena. *Organizace znalostí: klíčová témata*. Praha: Karolinum, 2017. 269 s. ISBN 978-80-246-3587-3
- ❖ GOLUB, Koraljka. *Subject access to information: an interdisciplinary approach*. Santa Barbara, California: Libraries Unlimited, 2015. 165 s. ISBN 978-1-61069-577-0.
- ❖ ROWLEY, Jennifer – HARTLEY, Richard. *Organizing knowledge: an introduction to managing access to information*. 4th ed. Farnham: Ashgate, 2008. 367 s. ISBN 978-0-7546-4431-6.
- ❖ ZENG, Marcia Lei a Jian QIN. *Metadata*. Second edition. London: Facet Publishing, 2016. ISBN 978-1-78330-052-5.
- ❖ KEYSER, Pierre de. *Indexing: from thesauri to the semantic web*. 1st pub. Oxford: Chandos, 2012. 249 s. ISBN 978-1-84334-292-2.
- ❖ ECO, Umberto. *Bludiště seznamů*. Praha: Argo, 2009. 408 s. ISBN 978-80-257-0164-5.
- ❖ ECO, Umberto. *Hledání dokonalého jazyka v evropské kultuře*. Praha: NLN, Nakladatelství Lidové noviny, 2001. 355 s. ISBN 80-7106-389-4.
- ❖ LAKOFF, George. *Ženy, oheň a nebezpečné věci. Co kategorie vypovídají o naší mysli*. Praha: Triáda, 2006. 655 s. ISBN 80-86138-78-X.

