

Téma 3 - Indexace

Indexace může znamenat:

- tvorbu indexu (indexového souboru)
- vyjadřování obsahu (věcný popis / zpracování) informací
⇒ **obsahová metadata**

1

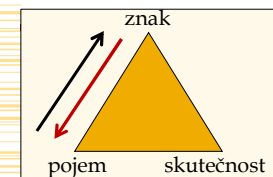
Indexace jako popis obsahu zdrojů

2

Proces indexace podle ISO 5963

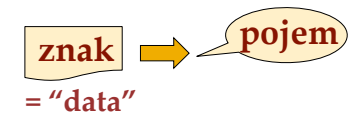
- 1) Analýza požadavku / dokumentu a určení jeho věcného **obsahu / předmětu**
- 2) Identifikace hlavních **pojmu** obsažených v předmětu
- 3) Vyjádření těchto pojmů **lexikálními jednotkami** selekčního jazyka

3



Proces indexace podle ISO 5963

- 1) Obsahová analýza



- 2) Vyjádření obsahu dokumentu nebo informačního dotazu prostřednictvím slov přirozeného jazyka nebo znaků umělého jazyka



4

(Obsahová) metadata jako komunikační infra- nebo hyper-struktura 1

infra = dole, vespod, pod
hyper = nad, přes, nadmíru



1)



2)



3)

ÚISK – OBA T03 5

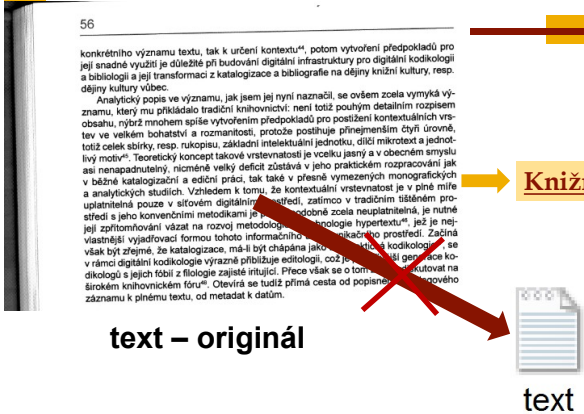
Obsahová analýza textu

56

konkrétního významu textu, tak k určení kontextu⁴⁴, potom vytvoření předpokladů pro její snadné využití je důležité při budování digitální infrastruktury pro digitální kodikologii a bibliologii a její transformaci z katalogizace a bibliografie na dějiny knižní kultury, resp. dějiny kultury vůbec.

Analytický popis ve významu, jak jsem jej nyní naznačil, se ovšem zcela vymyká významu, který mu přiřadalo tradiční knihovnictví: není totiž použitým detailním rozpisem textu ve velkém bohatství a rozmanitosti, protože postihuje přinejmenším čtyři úrovně, totiž celek sbírky, resp. rukopisu, základní intelektuální jednotku, dílčí mikrotext a jednotlivý motiv⁴⁵. Teoretický koncept takové vrstevnatosti je vcelku jasný a v obecném smyslu asi nenapadnutelný, nicméně velký deficit zůstává v jeho praktickém rozpracování jak v běžné katalogizaci a ediční práci, tak také v přesně vymezených monografických a analytických studiích. Vzhledem k tomu, že kontextuální vrstevnatost je v piné míře uplatnitelná pouze v síťovém digitálním prostředí, zatímco v tradičním tištěném prostředí a jeho konvenčních metodikami je neobdobně zcela neuplatnitelná, je nutné její zpřístupňování vázat na rozvoj metodologie a bibliologie hypertextu⁴⁶, jež je nejlépe vyjadřovací formou tohoto informačního a komunikativního prostředí. Základní podmínkou je však být zřejmé, že katalogizace, má-li být chápána jako „kodikologie“, se v rámci digitální kodikologie výrazně přibližuje edilogii, což je pro její generace kodikologů s jejich fóbií z filologie zajiště intuící. Přece však se o tom nemějí diskutovat na širokém knihovnickém fóru⁴⁷. Otevírá se tudíž přímá cesta od popisného katalogového záznamu k plnému textu, od metadat k datům.

Knižní šifra



text – originál

text

ÚISK – OBA T03 6

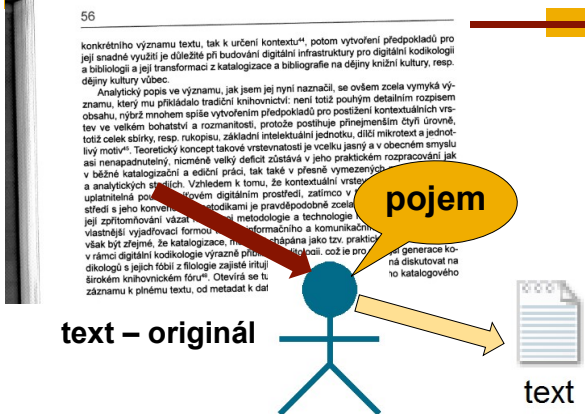
Obsahová analýza textu

56

konkrétního významu textu, tak k určení kontextu⁴⁴, potom vytvoření předpokladů pro její snadné využití je důležité při budování digitální infrastruktury pro digitální kodikologii a bibliologii a její transformaci z katalogizace a bibliografie na dějiny knižní kultury, resp. dějiny kultury vůbec.

Analytický popis ve významu, jak jsem jej nyní naznačil, se ovšem zcela vymyká významu, který mu přiřadalo tradiční knihovnictví: není totiž použitým detailním rozpisem textu ve velkém bohatství a rozmanitosti, protože postihuje přinejmenším čtyři úrovně, totiž celek sbírky, resp. rukopisu, základní intelektuální jednotku, dílčí mikrotext a jednotlivý motiv⁴⁵. Teoretický koncept takové vrstevnatosti je vcelku jasný a v obecném smyslu asi nenapadnutelný, nicméně velký deficit zůstává v jeho praktickém rozpracování jak v běžné katalogizaci a ediční práci, tak také v přesně vymezených monografických a analytických studiích. Vzhledem k tomu, že kontextuální vrstevnatost je v piné míře uplatnitelná pouze v síťovém digitálním prostředí, zatímco v tradičním tištěném prostředí a jeho konvenčních metodikami je neobdobně zcela neuplatnitelná, je nutné její zpřístupňování vázat na rozvoj metodologie a bibliologie hypertextu⁴⁶, jež je nejlépe vyjadřovací formou tohoto informačního a komunikativního prostředí. Základní podmínkou je však být zřejmé, že katalogizace, má-li být chápána jako tzv. praktická kodikologie, se v rámci digitální kodikologie výrazně přibližuje edilogii, což je pro její generace kodikologů s jejich fóbií z filologie zajiště intuící. Přece však se o tom nemějí diskutovat na širokém knihovnickém fóru⁴⁷. Otevírá se tudíž přímá cesta od popisného katalogového záznamu k plnému textu, od metadat k datům.

pojmem



text – originál

text

ÚISK – OBA T03 7

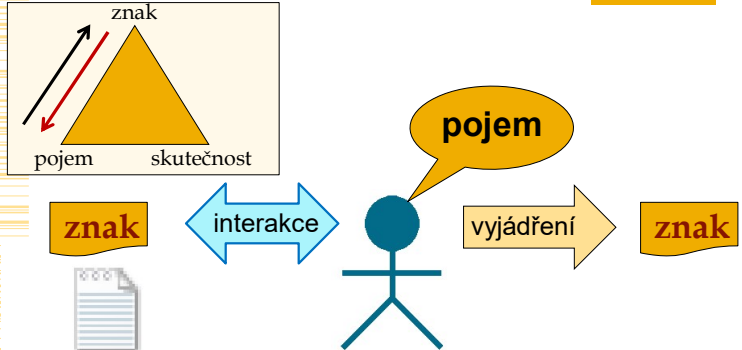
Indexace textu

56

konkrétního významu textu, tak k určení kontextu⁴⁴, potom vytvoření předpokladů pro její snadné využití je důležité při budování digitální infrastruktury pro digitální kodikologii a bibliologii a její transformaci z katalogizace a bibliografie na dějiny knižní kultury, resp. dějiny kultury vůbec.

Analytický popis ve významu, jak jsem jej nyní naznačil, se ovšem zcela vymyká významu, který mu přiřadalo tradiční knihovnictví: není totiž použitým detailním rozpisem textu ve velkém bohatství a rozmanitosti, protože postihuje přinejmenším čtyři úrovně, totiž celek sbírky, resp. rukopisu, základní intelektuální jednotku, dílčí mikrotext a jednotlivý motiv⁴⁵. Teoretický koncept takové vrstevnatosti je vcelku jasný a v obecném smyslu asi nenapadnutelný, nicméně velký deficit zůstává v jeho praktickém rozpracování jak v běžné katalogizaci a ediční práci, tak také v přesně vymezených monografických a analytických studiích. Vzhledem k tomu, že kontextuální vrstevnatost je v piné míře uplatnitelná pouze v síťovém digitálním prostředí, zatímco v tradičním tištěném prostředí a jeho konvenčních metodikami je neobdobně zcela neuplatnitelná, je nutné její zpřístupňování vázat na rozvoj metodologie a bibliologie hypertextu⁴⁶, jež je nejlépe vyjadřovací formou tohoto informačního a komunikativního prostředí. Základní podmínkou je však být zřejmé, že katalogizace, má-li být chápána jako tzv. praktická kodikologie, se v rámci digitální kodikologie výrazně přibližuje edilogii, což je pro její generace kodikologů s jejich fóbií z filologie zajiště intuící. Přece však se o tom nemějí diskutovat na širokém knihovnickém fóru⁴⁷. Otevírá se tudíž přímá cesta od popisného katalogového záznamu k plnému textu, od metadat k datům.

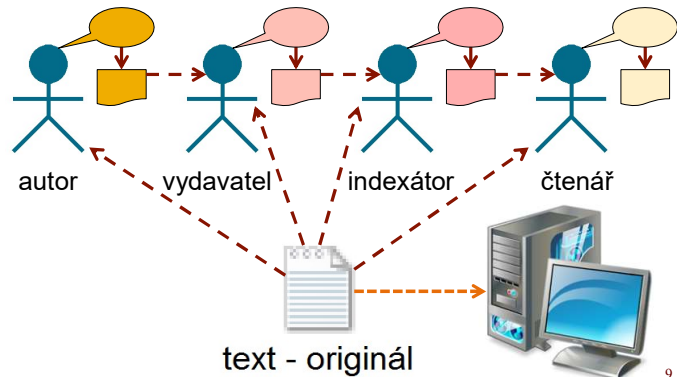
znak



text - originál

ÚISK – OBA T03 8

Faktory obsahové analýzy: subjekt, kontext, účel



9

Pragmatický aspekt indexace

„*warrant*“ = zdůvodněný, ospravedlněný, oprávněný

- literary warrant
- educational warrant
- user warrant
- organizational warrant
- business warrant

Příklad:

Kniha v TDKIV
Kniha v tezauru AGROVOC

„*bias*“ = zaujetí, zkreslení, předpojatost, předsudek,
subjektivní ovlivnění

ÚISK – OBA T03

10

10

Indexace složených témat

Vyjádření syntagmatických vztahů mezi dílčími tématy:

- (1) Fasetová analýza
- (2) „Závislostní“ analýza
- (3) Předmětová hesla

ÚISK – OBA T03

11

11

Fasetová obsahová analýza = faseta jako kategorie

Obecné fasety/kategorie
aplikovatelné na velké množství typů obsahu

- | | |
|--------------------|-----------------------|
| ■ věc (co) | ■ zpracování potravin |
| ■ agent (kdo) | ■ vítěz závodu |
| ■ proces (činnost) | ■ heterogenní síť |
| ■ událost | ■ porcelánová soška |
| ■ materiál | ■ start raketoplánu |
| ■ vlastnost (jaký) | ■ doprava v Praze |
| ■ místo (kde) | ■ pondělní výuka |
| ■ čas (kdy) | ■ karlovarský kaolin |
| | ■ mistrovství Evropy |

ÚISK – OBA T03

12

12

Struktura složených témat

(1) Souřadná

symetrická

Sever a Jih, českomoravský, Jü a Hele,
hluchoněmý

Klasifikace a systémová analýza

(2) Podřadná

asymetrická

Jižní Čechy, moravské uzené, Alexandre Dumas
mladší, prvotisk

Etické faktory informačních interakcí člověka

ÚISK – OBA T03

13

13

„Závislostní“ analýza složených témat vytvořených podřadným skládáním

1. Určete významové jádro (*term focus*)
2. Určete rozlišující prvky (*term difference*)
3. Určete hledisko (*aspekt*), ev. žánr/formu či uživatelské určení – je-li aplikovatelné
4. Zformulujte termíny pro jednotlivé prvky

Příklad

Restaurování šelakových gramofonových desek

významové jádro	rozlišující prvek	hledisko
gramofonová deska	šelak (materiál)	restaurování

ÚISK – OBA T03

14

14

„Závislostní“ analýza je syntagmatická

1. Synchronní a diachronní **výzkum** češtiny z hlediska strukturálního, kognitivního a funkčního
2. Financování základního **výzkumu** v AV ČR
3. Péče o talentovanou populaci v období docházky do mateřské a základní školy: průběžná **výzkumná** zpráva
4. Informační systém **výzkumu**, vývoje a inovací

významové jádro	rozlišující prvek	hledisko

ÚISK – OBA T03

15

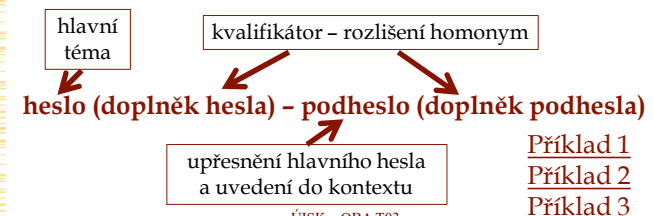
15

Předmětová hesla

ČSN 01 0188. *Tvorba předmětových hesel*. 1983

pravidla:

- nejužšího pojmu
- slovních spojení
- jazykové úpravy



ÚISK – OBA T03

16

16

Typy podhesel a pořadí jejich zápisu

1.	věcné upřesnění / zúžení významu hesla	<i>Akvizice ve veřejných knihovnách</i>	Knihovny veřejné - akvizice
2.	pojem ve vztahu k heslu (např. další téma)	<i>Veřejné knihovny a krajanské spolky</i>	Knihovny veřejné - spolky krajanské
3.	vztah hesla a podhesla typu 2	<i>Spolupráce veřejných knihoven a krajanských spolků</i>	Knihovny veřejné - spolky krajanské - spolupráce
4.	zeměpisné určení	<i>Spolupráce českých veřejných knihoven a krajanských spolků</i>	Knihovny veřejné - spolky krajanské - spolupráce - Česko
5.	časové určení	<i>Spolupráce českých veřejných knihoven a krajanských spolků ve 21. století</i>	Knihovny veřejné - spolky krajanské - spolupráce - Česko - 21. století
6.	forma/žánr	<i>Výzkum spolupráce českých veřejných knihoven a krajanských spolků ve 21. století</i>	Knihovny veřejné - spolky krajanské - spolupráce - Česko - 21. století - studie

ÚISK – OBA T03

17

17

Proč potřebujeme syntax

Samotné pojmy (někdy) nestačí

Studie pojednává o **hudební sbírce kantorského rodu Strachotů z Panenského Týnce** uložené v **hudebním oddělení Národní knihovny České republiky**.

Knihovna
knihovnická revue



ŠEDIVÁ, Eliška. Hudební sbírka kantorské rodiny Strachotů z Panenského Týnce ve fondu hudebního oddělení Národní knihovny České republiky. *Knihovna: knihovnická revue*, 2023, 34(1), s. 5-27. ISSN 1801-3252. Dostupné také z: <https://knihovnavuevue.nkp.cz/archiv/2023-1/recenzovane-prispevky/hudebni-sbirka-kantorske>.

ÚISK – OBA T03

18

18

Koordinace

souřadné spojení (syntagmatické)

využití **více lexikálních jednotek** k vyjádření **složených témat** (typ kolokace)

prekoordinace

vztahy definovány už při tvorbě řízeného slovníku nebo při indexaci (na vstupu, apriori, „paradigmaticky“)
zvyšuje **přesnost**

postkoordinace

vztahy definovány až při formulaci požadavku/dotazu (na výstupu, a posteriori, syntagmaticky)
zvyšuje **úplnost**

ÚISK – OBA T03

19

19

prekoordinace

popis dokumentu

*hodnocení profesorů
studenty* ●

popis dotazu

*hodnocení profesorů
studenty* ●

postkoordinace

popis dokumentu

hodnocení ○
profesoři ○
studenti ○

popis dotazu

hodnocení
AND
profesoři
AND
studenti

ÚISK – OBA T03

20

20

	prekoordinace	postkoordinace
klady	<ul style="list-style-type: none"> ■ přesné vyjádření tématu dokumentu (i velmi specifické vztahy) 	<ul style="list-style-type: none"> ■ možnost libovolného skládání pojmů
záporny	<ul style="list-style-type: none"> ■ obtížná formulace dotazu běžným uživatelem ■ nevyhovuje interdisciplinárním dotazům 	<ul style="list-style-type: none"> ■ omezení typů vztahů, jež lze použít

ÚISK – OBA T03 21

21

Prekoordinace vs. postkordinace

Vliv a účinky volného času na jednotlivce v České republice a na Slovensku v 21. století

MDT

379.8-042.3-059.1"20"(437.3+437.6)

1 notační znak = 1 téma
(prekoordinace)

Tezaurus EUROVOC

volný čas
Česká republika
Slovensko
sociální analýza
21. století

1 téma = kombinace preferovaných termínů
(postkoordinace)

ÚISK – OBA T03

22

Prekoordinace v MDT

Vliv a účinky volného času na jednotlivce v České republice a na Slovensku v 21. století

379.8-042.3-059.1"20"(437.3+437.6)

379.8	Volný čas. Dovolena, prázdniny	hlavní znak
-042.3	Vliv	všeobecný pomocný znak vztahů, procesů a operací (postupů)
-059.1	Osoby jednotlivě. Jednotlivé osoby	všeobecný pomocný znak osob a osobních charakteristik
"20"	21. století	všeobecný pomocný znak času
(437.3)	Česká republika	všeobecný pomocný znak místa
+	přifažení	všeobecný pomocný symbol
(437.6)	Slovensko. Slovenská republika	všeobecný pomocný znak místa

-3

23

Postkoordinace v LISS

- Looking Ahead: Incorporating AI in MLIS Competencies.**

By: Ghosh, Souvik; McCoy, Denise. *Student Research Journal (SRJ)*, May2024, Vol. 14 Issue 2, p28-42, 15p

Témata: Digital literacy; Information organization; Democratization; Educational programs; Internet access; Educational quality

Akademický časopis Plný text PDF (1.8MB) Find author's articles in Web of Science
- Knowledge Organization and Representation under the AI Lens.**

By: Qin, Jian. *Journal of Data & Information Science*, Feb 2020, Vol. 5 Issue 1, p3-17, 15p; DOI: 10.2478/jdis-2020-0002

Témata: Knowledge representation (information theory); Artificial intelligence; Information science; Comparative literature; Case studies

Akademický časopis Plný text HTML Plný text PDF (3MB) Find author's articles in Web of Science
- Knowledge Organization Systems: A Network for AI with Helping Interdisciplinary Vocabulary Engineering**

By: Greenberg, Jane; Zhao, Xintong; Monselise, Michal; Grabus, Sam; Boone, Joan. *Cataloging & Classification Quarterly*, 2021, Vol. 59 Issue 1, p1-12, 12p; DOI: 10.1080/00090747.2021.1911111

Témata: Knowledge management; Natural language processing; Machine learning; Artificial intelligence; Digital resources in library science

Akademický časopis Odkazovaný plný text Find author's articles in Web of Science

24

Postkoordinace v JSTOR

□ Knowledge Organization and Text Organization

Suzanne M. Mannes, Walter Kintsch
Cognition and Instruction, Vol. 4, No. 2 (1987), pp. 91-115
 JOURNAL ARTICLE

Topics: [Full text databases](#), [Advance organizers](#), [Inference](#), [Problem solving](#), [Paraphrase](#), [Bacteria](#),
[Learning](#), [Memory retrieval](#), [Statistical significance](#)

□ Effects of Knowledge Organization on Task Performance

Bat-Sheva Eylon, F. Reif
Cognition and Instruction, Vol. 1, No. 1 (Winter, 1984), pp. 5-44
 JOURNAL ARTICLE

Topics: [Physics](#), [History](#), [Problem solving](#), [Experiment design](#), [Performance tests](#), [Instructional materials](#),
[Knowledge acquisition](#), [Educational psychology](#), [Cognitive models](#), [Memory](#)

:5

25

Komentář k úkolu 5

Systemová analýza:

- **Třída/kategorie** (substantivum)
- **Vlastnost – atribut** (adjektivum), **vztah** (verbum)
- **Prvek/instance** (proprium)

ÚISK – OBA T03

26

26

Komentář k úkolu 5 Při obsahové analýze pozor na:

- **přídavná jména** (adjektiva)
moderní (≠ *modernita*), *umělý* (*umělý život*), *střední*, *digitální*
 často nemají samostatný význam
- **slovesa**
použití, *využití*, *vytvoření*, *význam*, *důvod*, *souvislost*, *dopad*,
vliv, *zlepšení*, *důsledek*
 spíše vztah (predikát), nikoli samostatná třída/téma
- **„prázdná“, příliš obecná slova**
téma, *problém*, *otázka*, *aspekt*, *problematika*
- **vlastní jména** (spíše instance než třída/pojem)
Havířov, *Noc s Andersenem*, *Inforum*, *70 let*

ÚISK – OBA T03

27

27

Komentář k úkolu 5 Jsou všechny složené pojmy zpětně rozložitelné?

- | | |
|------------------------|-----------------------------|
| ■ teorie informace | ■ informační potřeby |
| ■ autonomní vozidlo | ■ bibliografická data |
| ■ digitalizační plán | ■ destruktivní digitalizace |
| ■ statistika výpůjček | ■ absenční výpůjčky |
| ■ význam biodiverzity | ■ mimoškolní vzdělávání |
| ■ slovní druhy | ■ nová média |
| ■ obsahová analýza | ■ moderní tanec |
| ■ šachový turnaj | ■ neočekávaná výhoda |
| ■ terminologická norma | ■ mezinárodní standard |

ÚISK – OBA T03

28

28

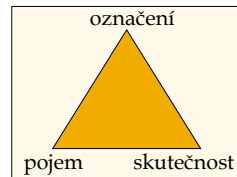
Komentář k úkolu 6

Složený pojem ≠ víceslovný výraz

- novostavba
- bohemikum
- jezdec
- lyžař
- vydavatel
- origami

- jak vzniká
- vybrané mapové sbírky
- rozvoj digitální gramotnosti
- některé problémy
- výsledky průzkumu
- zlatá cesta

slovo ≠ pojem ≠ věc



ÚISK – OBA T03

29

29

Komentář k úkolu 6

Kdy je to vlastnost a kdy věc?

- abecední řazení
- absenční výpůjčka
- bibliografický formát
- meruňkový koláč
- meruňková barva
- klasifikační systém
- krokodýlí kůže
- veselý výlet
- zkouškové období

Připomenutí z předmětu KSA

Vztah substantivum –
adjektivum je vztahem
třída – vlastnost.

Vlastnost =

- a) atribut
- b) vztah (k jiné třídě)
⇒ **relační adjektiva**

ÚISK – OBA T03

30

30

Komentář k úkolu 6

Co je významové jádro

– proces/vlastnost nebo věc (co závisí na čem)?

- dezinfekce vody
- tvorba tezauru
- šíření informací
- relevance informací
- digitalizační pracoviště
- registr digitalizace
- katalogizace hudebnin
- interakce člověk–počítač

Zlatá cesta otevřeného přístupu

ÚISK – OBA T03

31

31

Komentář k úkolu 6

Priorita substantiva nebo adjektiva?

- potřeby informační
- vozidlo autonomní
- plán digitalizační
- výpůjčky absenční
- druhy slovní
- analýza obsahová
- turnaj šachový
- norma terminologická
- vzdělávání mimoškolní
- data bibliografická
- informační potřeby
- autonomní vozidlo
- digitalizační plán
- absenční výpůjčky
- slovní druhy
- obsahová analýza
- šachový turnaj
- terminologická norma
- mimoškolní vzdělávání
- bibliografická data

ÚISK – OBA T03

32

32