

Téma 4 Systém organizace znalostí

- Jako model metadat (⇒ pojmová indexace)
- Jako řízený slovník (⇒ vyjádření pojmů, notace)
- Jako informační zdroj / dokument

ÚISK – OBA T04 1

1

Výsledek (produkty) procesu organizace znalostí

- **struktura** = organizovaný zdroj
 - dokument, záznam, soubor, digitální objekt...
 - fond, kolekce, sbírka, archiv, web site, digitální knihovna, repozitář, databáze, datová sada (dataset), znalostní báze, znalostní graf... primární informace / „data“
- **metadata** / organizovaný soubor **metadat**
 - katalog, index, registr... sekundární informace
- **pomůcky** pro tvorbu metadat = nástroje a metodiky
 - a) pravidla, metodiky, techniky, příklady dobré praxe...
 - b) metadatová schémata / modely systémy organizace znalostí
 - c) slovníky – řízené slovníky

ÚISK – OBA T04 2

2

Systém organizace znalostí (selekční jazyk)

Extenzionální definice (výčetem):

Gail Hodgeová

Termín systém organizace znalostí je určen k označení všech typů schémat pro organizaci informací a podporu znalostního managementu. Systémy organizace znalostí zahrnují **klasifikační a kategorizační schémata**, jež organizují materiál na všeobecné úrovni, **předmětová hesla**, jež umožňují detailnější přístup, a **soubory autorit**, jež slouží k řízení variantních verzí klíčových informací, jako například geografických nebo osobních jmen...zahrnují rovněž vysoce strukturované slovníky, jako jsou **tezaury**, a méně tradiční schémata jako **sémantické sítě** a **ontologie**.

ÚISK – OBA T04 3

3

Typy systémů organizace znalostí podle typu vztahů (struktury)

kategorie	typ	anglicky	
pojmové sítě síťová struktura	16 (terminologická) ontologie	(terminological) ontology	G F + odvozování
	15 sémantická síť	semantic network	F E + asociace
	14 tezaurus	thesaurus	
	13 klasifikační schéma / systém	classification scheme	
klasifikace stromová struktura	12 taxonomie (biologie)	taxonomy	E D + hierarchie
	11 kategorizační schéma / systém	categorization scheme	
	10 předmětový heslář	subject heading scheme	
slovníky lineární struktura	9 seznam jmenných autorit	name authority list/file	D C + preferované termíny
	8 slovník zeměpisných názvů	gazetteer	
	7 seznam synonym	synonym ring, synset	C B + ekvivalence
	6 terminologický slovník	terminology	
	5 (výkladový) slovník, glosář	glossary	B A + definice
	4 číselník (nomenklatura, seznam, kódovník)	list, code list	
	3 (řízený) slovník	(controlled) vocabulary, dictionary	A seznam slov
2 negativní slovník (stop slova, zakázaná slova)	stopwords, exclusion list		
1 nabídkový seznam (menu)	pick list		

↑
sémantická síla

ÚISK – OBA T04 4

4

Klasifikační schémata, tezaury a ontologie

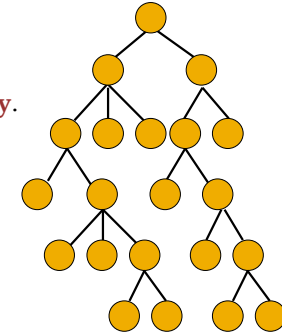
5

Klasifikační schémata, tezaury a ontologie

Systémy organizace znalostí uspořádané tak, že explicitně zachycují **paradigmatické** (sémantické) **vztahy** mezi **pojmy**.

Nástroj vyjádření vztahů:
notace

Funkce:
indexace informačních zdrojů



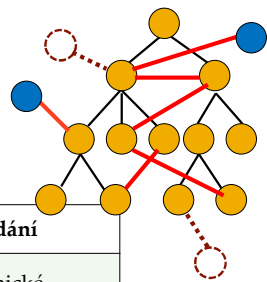
ÚISK – OBA T04

6

6

Klasifikační schémata, tezaury a ontologie

název základní jednotky	●
klasifikační schéma, ontologie	třída
tezaurus	termín
Wikidata	položka



základní paradigmatické vztahy		uspořádání
klasifikační schéma, ontologie	hierarchie	hierarchické
tezaurus	ekvivalence	abecední (lineární)
Wikidata	asociace	síťové

ÚISK – OBA T04

7

7

MDT - slovník

- **hlavní znak**
 - jednoduchý znak
 - rozvinutý znak (hlavní + pomocný znak)
 - složený znak (>1 hlavní znak)
- **pomocný znak**
 - všeobecný
 - samostatný
 - závislý
 - zvláštní

ÚISK – OBA T04

8

8

MDT - syntax spojovací symboly

- | | | |
|-----|---------------------|----------------|
| + | přiřazení | } celek - část |
| / | rozšíření | |
| : | vztah (symetrický) | } asociace |
| :: | nezaměnitelný vztah | |
| [] | podřazení | |

ÚISK - OBA T04

9

9

Fasety v Mezinárodním desetinném třídění

Universal Decimal Classification

summary

ZÁKLADNÍ SCHEMA SYMBOLY POMOCNÉ ZNAKY 0 1 2 3 4 5 6 7 8

zobrazit vše | skrýt vše

ZÁKLADNÍ SCHEMA

POMOCNÉ TABULKY

VŠEOBECNÉ POMOCNÉ ZNAKY A JEJICH ČLENĚNÍ

COMMON AUXILIARY NUMBERS

[...] Všeobecné pomocné znaky jazyka. Tabulka 1c.

[0..] Všeobecné pomocné znaky formy. Tabulka 1d.

[1/9] Všeobecné pomocné znaky místa. Tabulka 1e

[...] Všeobecné pomocné znaky plemen, etnických skupin a národností. Tabulka 1f.

[...] Všeobecné pomocné znaky času. Tabulka 1g

[...] Všeobecné pomocné znaky obecných charakteristik: vlastnosti, materiály, procesy a operace, osoby

Hlavní tabulky

0 VĚDA A VĚDĚNÍ, ORGANIZACE, VĚDA O POČÍTAČÍCH, VÝPOČETNÍ TECHNIKA, I

1 FILOZOFIE, PSYCHOLOGIE

2 NÁBOŽENSTVÍ, TEOLOGIE

3 SOCIÁLNÍ VĚDY

4 MATEMATIKA, PŘÍRODNÍ VĚDY

5 APLIKOVANÉ VĚDY, LÉKAŘSTVÍ, TECHNIKA

6 UMĚNÍ, REKREACE, ZÁBAVA, SPORT

7 JAZYK, LINGVISTIKA, LITERATURA

8 GEOGRAFIE, BIOGRAFICKÉ STUDIE, DĚJINY

ÚISK - OBA T04

10

10

Formální fasety

Obsahová faseta

MDT - všeobecné pomocné znaky

- = jazyka
- (0..) formy dokumentu
- () místa
- ":" času
- (=) etnický
- 0 obecné charakteristiky (vlastnosti, materiály, vztahy, procesy a operace, osoby)

ÚISK - OBA T04

11

11

MDT - zvláštní pomocné znaky

-1/-9 analytické

.01/.09 opakující se charakteristiky

'1/'9 syntetické

ÚISK - OBA T04

12

12

Plně fasetové třídy v MDT

- 2 Náboženství. Teologie
- 7 Umění. Rekreace. Zábava. Sport
- 8 Jazyk. Lingvistika. Literatura
- 93/94 Dějiny
- ...a další

ÚISK – OBA T04

13

13

MDT – způsob vyjádření složených témat

Fasetový

1 téma charakterizované z více hledisek

kombinace znaků ⇒ rozvinuté znaky

hlavní znak a všeobecný pomocný znak

hlavní znak a zvláštní pomocný znak

75(031)

Encyklopedie malířství

75.034.7

Barokní malířství

Syntaktický

více témat

koordinace znaků pomocí

spojovacích symbolů

⇒ složené znaky

75+73

Malířství a sochařství

(031+038)

Encyklopedie a slovníky

159.9:7

Psychologie umění

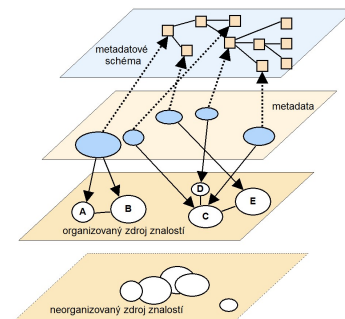
ÚISK – OBA T04

14

14

System organizace znalostí

- Jako model metadat (⇒ pojmová indexace)
- Jako řízený slovník (⇒ vyjádření pojmů, notace)
- Jako informační zdroj / dokument



ÚISK – OBA T04

15

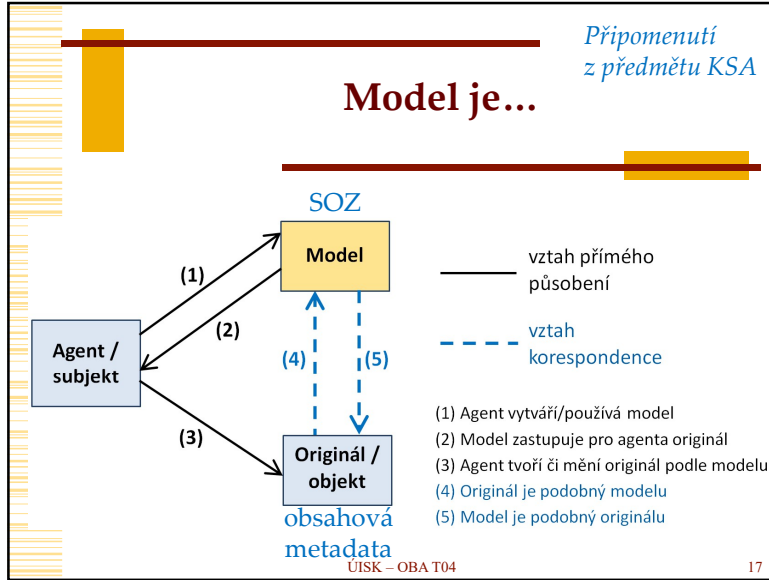
15

System organizace znalostí jako model metadat

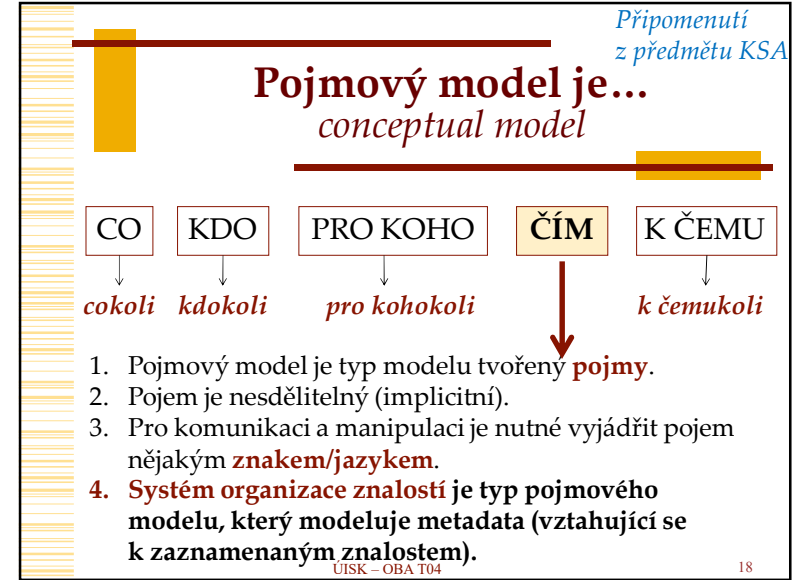
ÚISK – OBA T04

16

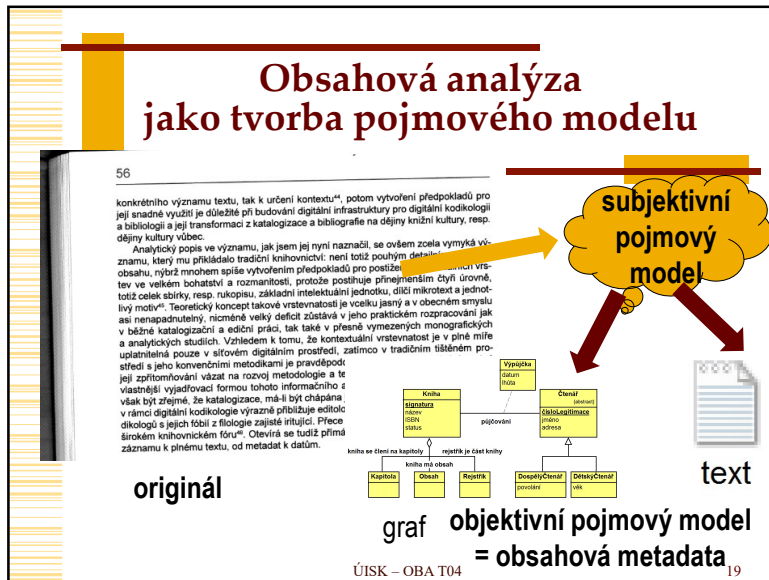
16



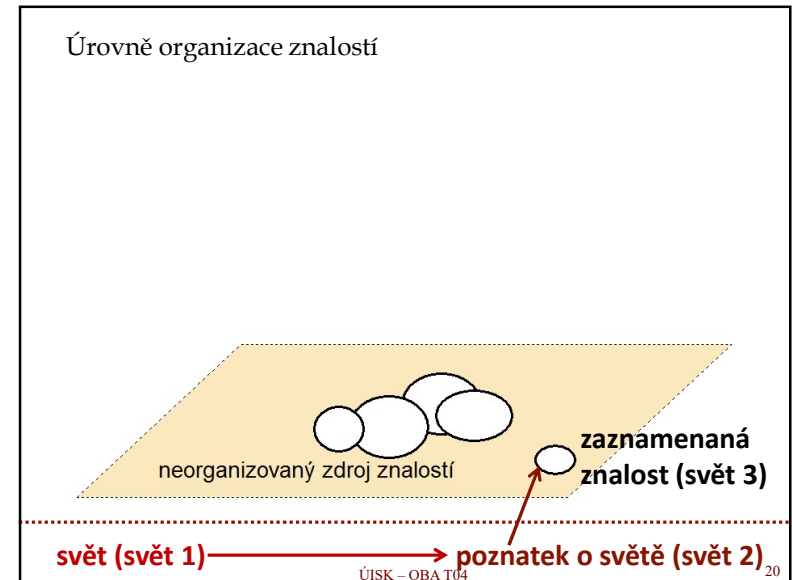
17



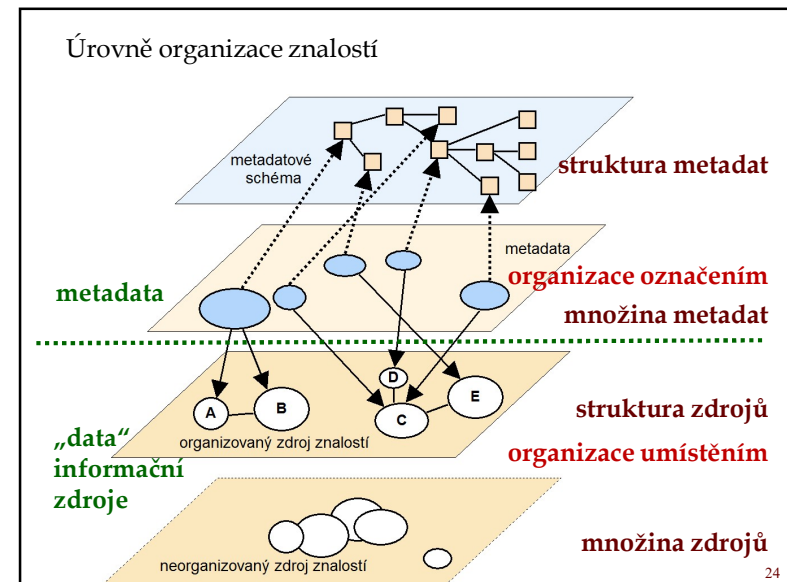
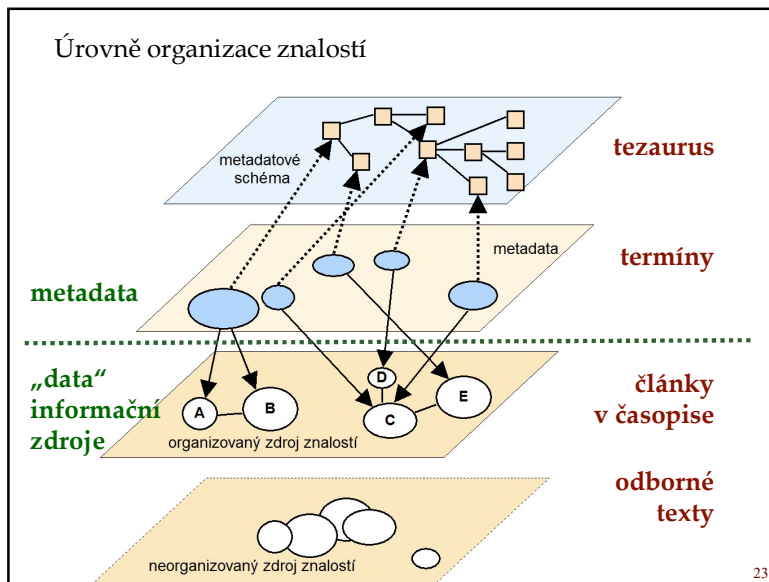
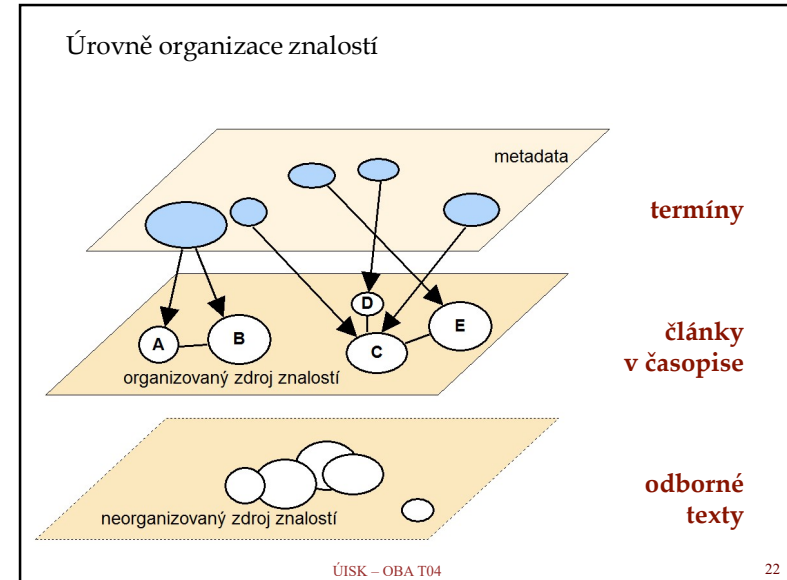
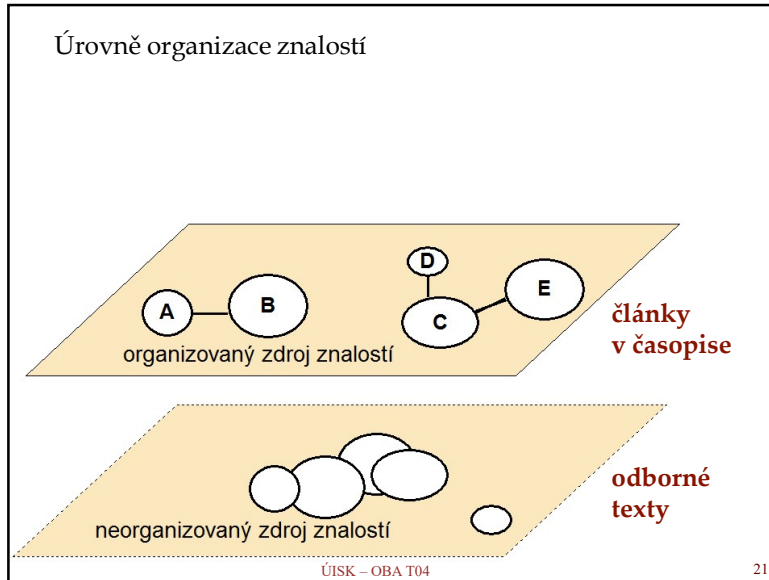
18



19



20



21

22

23

24

Účel systému organizace znalostí

Pojmová indexace zdrojů v určené doméně

Odvozování z modelu:
Témata zdrojů A a E spolu souvisejí.

The diagram illustrates the process of organizing knowledge. It starts with 'neorganizovaný zdroj znalosti' (unorganized knowledge source) at the bottom, represented by scattered circles. These are organized into 'organizovaný zdroj znalosti' (organized knowledge source) in the middle, where circles are grouped and labeled A through E. This process is supported by 'metadata' and a 'metadatové schéma' (metadata schema) at the top, labeled 'SOZ'. Arrows indicate the flow of information from the unorganized to the organized state, and from the organized state to the metadata schema.

ÚISK – OBA T04 25

25

Paradigmatické vztahy v systémech organizace znalostí

ÚISK – OBA T04 26

26

Ekvivalence, hierarchie a asociace v tezauru Agrovoc – textové vyjádření

Připomenutí z předmětu KSA

PREFERRED TERM	① změna klimatu 🌡️
DEFINITION	① As alterações climáticas são mudanças do clima que ocorrem a nível global. Segundo o "Painel Intergovernamental sobre Mudanças Climáticas" (IPCC, 2001) as alterações climáticas resultam de uma variação estatisticamente significativa (no mínimo mais de 30 anos) no estado médio do clima ou na sua variabilidade, sem identificação absoluta das causas dessa mudança. (pt)
BROADER CONCEPT	přírodní jevy (cs) životní prostředí (cs)
NARROWER CONCEPTS	anthropogenic climate change (en) globální oteplení (cs)
RELATED CONCEPTS	adaptace na změnu klimatu (cs) climate system (en) zmírňení dopadů změny klimatu (cs)
ENTRY TERMS	③ změny klimatu (cs)
INFLUENCES	klima (cs) narušování ozonové vrstvy (cs)
IS INFLUENCED BY OR DEPENDS ON	obohacení kyslíčkem uhlíčitým (cs) skleníkové plyny (cs)
IS OBJECT OF ACTIVITY	UNFCCC (cs)

ÚISK – OBA T04 27

27

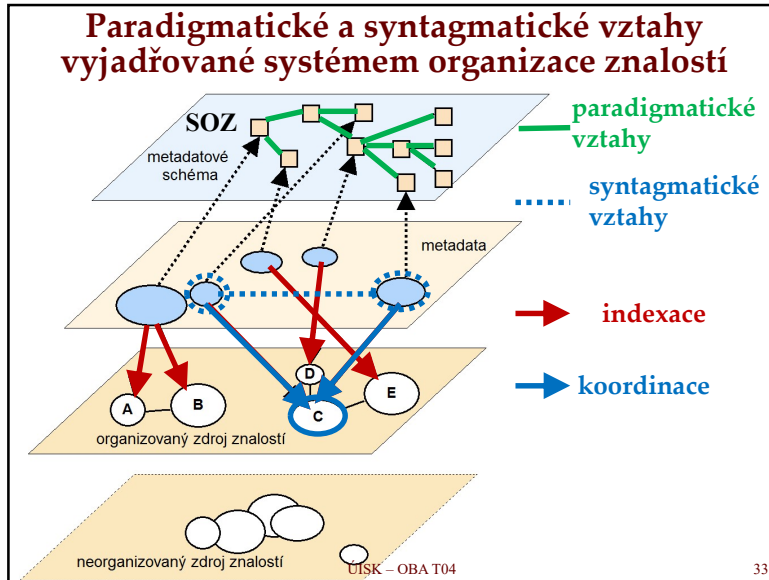
Ekvivalence, hierarchie a asociace v tezauru Agrovoc – grafické vyjádření

Připomenutí z předmětu KSA

The diagram shows 'změna klimatu' (climate change) as the central concept. It is influenced by 'obohacení kyslíčkem uhlíčitým' (CO2 enrichment), 'skleníkové plyny' (greenhouse gases), and 'UNFCCC'. It is the object of activity for 'anthropogenic climate change (en)' and 'globální oteplení' (global warming). 'změna klimatu' influences 'adaptace na změnu klimatu' (climate change adaptation), 'zmírňení dopadů změny klimatu' (mitigation of climate change impacts), and 'klimatický systém (climate system)'. It also influences 'narušování ozonové vrstvy' (ozone layer depletion) and 'klima' (climate). The diagram also shows 'přírodní jevy' (natural phenomena) and 'životní prostředí' (environment) as related concepts. A legend on the right explains the relationship types: (genetic) hierarchy (solid arrow), symmetric association (double-headed arrow), asymmetric association (arrow with open head), and equivalence (dashed arrow).

ÚISK – OBA T04 28

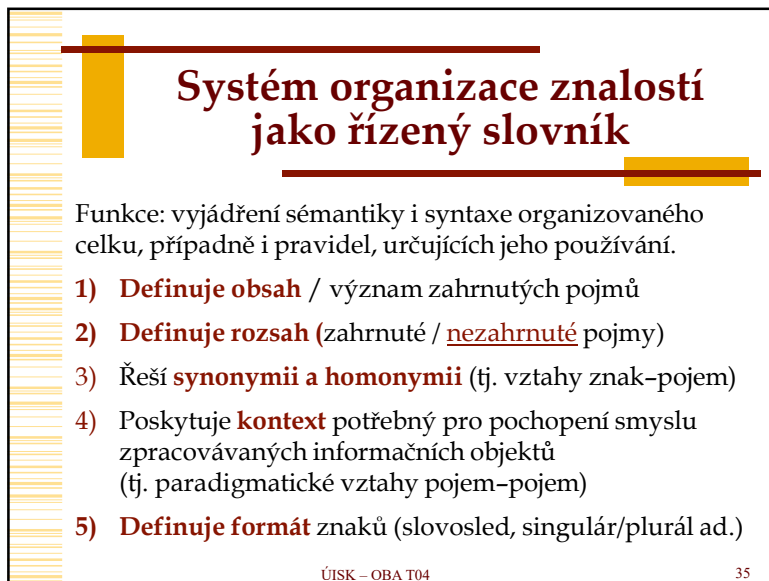
28



33



34



35



36

Proč potřebujeme slovník



Disk z Faistu
1700 př. n. l.

Tab. 01.01 Vybrané demografické údaje v České republice			
	1989	1990	1991
Počet obyvatel k 31. 12. (v tis. osob)			
muži	10 362	10 364	10 313
ženy	5 036	5 037	5 006
	5 326	5 327	5 307

ÚISK – OBA T04
37

37

Podmínka komunikace – jednotný kód (tj. slovník)




ÚISK – OBA T04


38


38

Sémantická interoperabilita

Kolik je hodin?
Который сейчас час? Wie spät ist es?
What time is it?



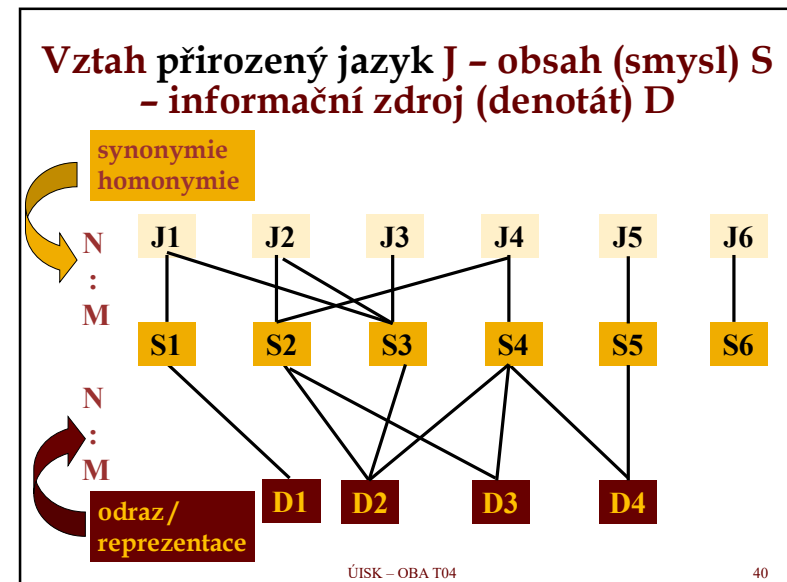




Без четверти семь.
 Es ist Viertel vor sieben.
 Je tři čtvrtě na sedm. Je ¾ na 7.
 It's a quarter to seven.
 18:45

ÚISK – OBA T04
39

39



40

Název **Organizace znalostí: klíčová témata / Helena Kučerová**

Kučerová, Helena, 1954-
Organizace znalostí: klíčová témata / Helena Kučerová. -- První vydání. -- Praha : Univerzita Karlova, nakladatelství Karolinum, 2017. -- 269 stran : grafy, tabulky
ISBN 978-80-246-3587-3

Typ obsahu: text
Typ média: bez média
Typ nosiče: svazek

Anglické resumé
-- Obsahuje bibliografické odkazy, bibliografii a rejstříky

* informace
* informační zdroje
* organizace informací
* dovednosti a znalosti

Témata: REF MESH deskriptor informační věda
General; Kno sémantika
eKniha data mining
management znalostí

3. Organizace znalostí Publikací typ monografie
Kučerová učebnice vysokých škol

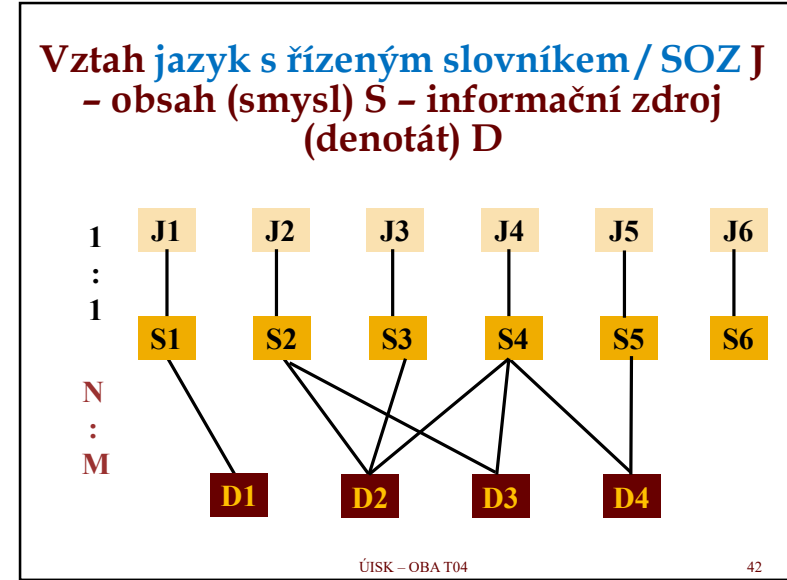
By: Kučerová, He NLK obory knihovnictví, informační věda a muzeologie
Czech, Databáze Konspekt 00

Témata: dovednosti a znalosti, informace, informační management, informační zdroje, organizace informací

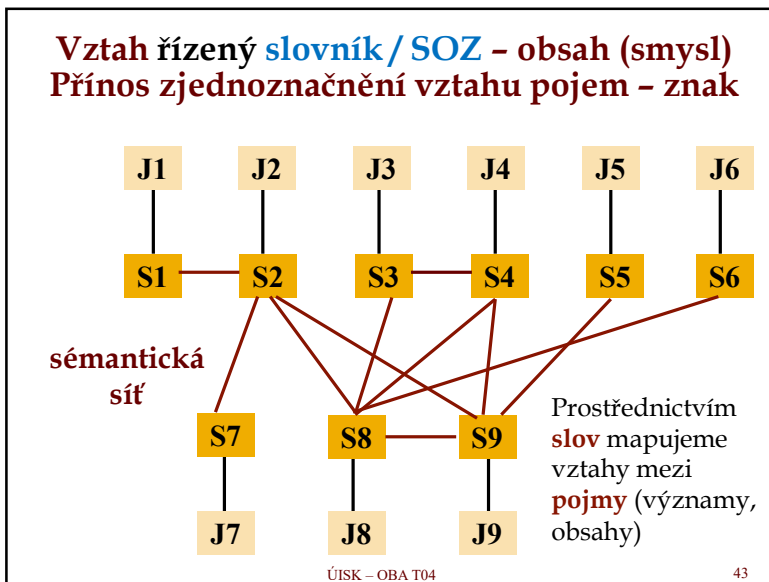
2017. 1 Ebook

ÚISK – OBA T04 41

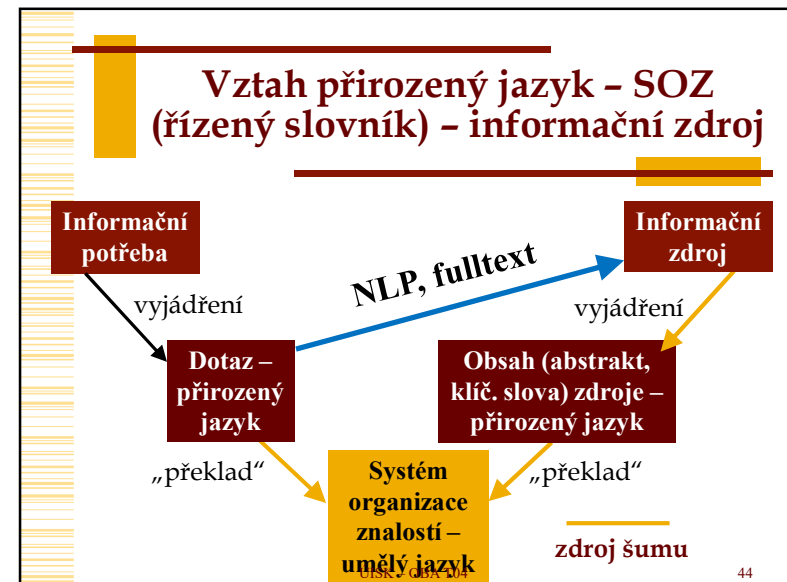
41



42

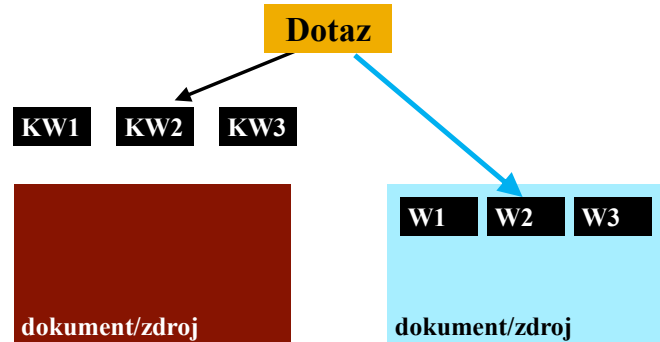


43



44

Přístup k textům s obsahovými metadaty a bez metadat



ÚISK – OBA T04

45

45

Přístup s obsahovými metadaty - pro a proti

- + vyřešená synonymie a homonymie
- + nezávislost na slovníku autora dokumentu (v metadatech je možné popsat obsah i slovy, jež nejsou v dokumentu)
- + k dispozici je model (abstrakce) obsahu, zbavený detailů
- + umožňuje pojmové vyhledávání
- + možnost interpretace, hodnocení
- vyžaduje intelektuálně předzpracovat dokumenty i dotazy
- v bodech zpracování dochází k šumu

ÚISK – OBA T04

46

46

Přístup bez obsahových metadat - pro a proti

- + dokumenty ani dotazy není třeba předem intelektuálně zpracovat
- + ubývá bodů zpracování (tj. zdrojů šumu)
- závislost na slovníku autora dokumentu
- problémy s přirozeným jazykem (synonymie, homonymie, vágnost)

ÚISK – OBA T04

47

47

Jak dostat to nejlepší z obou světů?

Přístup bez metadat:

- Místo porovnávání stručného textu dotazu s rozsáhlým textem dokumentu – statistické porovnávání (celého) textu relevantního zdroje s texty v prohledávaném souboru (např. při detekci spamu)

Přístup s metadaty:

- Místo intelektuálního vytváření metadat – automatické generování metadat, automatická indexace

ÚISK – OBA T04

48

48