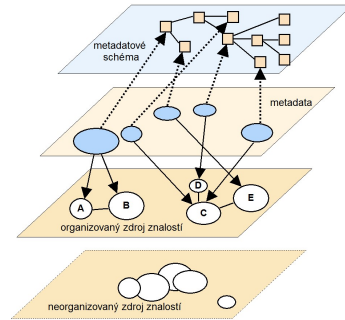


System organizace znalostí

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



ÚISK – OBA T04

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“**
- **metadát** / organizovaný soubor **metadat**
 - *catalog, index, registr...* **sekundární informace**
- **pomůcky** pro tvorbu metadat = nástroje a metodiky
 - pravidla, metodiky, standardy*
 - metadátová schémata / modely
 - slovníky – řízené slovníky, **systemy organizace znalostí**

ÚISK – OBA T04

2

Definice systému organizace znalostí (SOZ)

Funkce: podpora procesů organizace znalostí.

Obsah: Schéma, modelující strukturu (tj. prvky a vzájemné vztahy) organizované množiny znalostí.

Základní strukturální prvek: **pojem**.

⇒ **SOZ jako systém / model (metadat)**

Forma: Jádrem fyzické reprezentace každého systému organizace znalostí je **slovník**, tj. formální vyjádření pojmů.

Základní strukturální prvek: **termín**.

⇒ **SOZ jako (řízený) slovník**

ÚISK – OBA T04

3

System organizace znalostí (selekční jazyk)



Gail Hodgeová

Termín *system organizace znalostí* je určen k označení všech typů schémat pro organizaci informací a podporu znalostního managementu. Systemy 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

4

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	
klasifikace stromová struktura	13 klasifikační schéma / systém	classification scheme	
	12 taxonomie (biologie)	taxonomy	E D + hierarchie
	11 kategorizační schéma / systém	categorization scheme	
slovníky lineární struktura	10 předmětový heslář	subject heading scheme	
	9 seznam jmených 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	
	2 negativní slovník (stop slova, zakázaná slova)	stopwords, exclusion list	A seznam slov
1 nabídkový seznam (menu)	pick list		

↑
sémantická síla

ÚISK – OBA T04 5

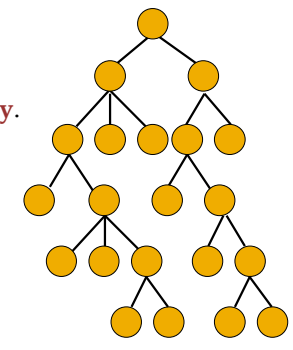
5

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

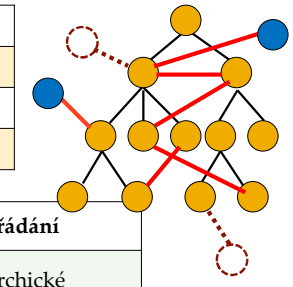


ÚISK – OBA T04 7

7

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 8

8

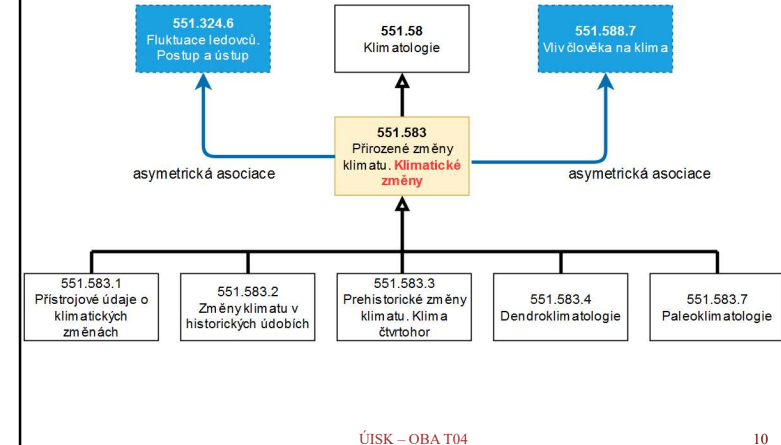
Hierarchie a asociace v MDT - textové vyjádření

↑ 551.58	Klimatologie
✍ 551.583	Přirozené změny klimatu. Klimatické změny ^
	⇒ 551.324.6 Fluktuační ledovců. Postup a ústup
	⇒ 551.588.7 Vliv člověka na klima
✍ 551.583.1	Přístrojové údaje o klimatických změnách
✍ 551.583.13	Periodické změny
✍ 551.583.15	Neperiodické změny
✍ 551.583.16	Klimatické extrémy
✍ 551.583.2	Změny klimatu v historických údobích
✍ 551.583.3	Prehistorické změny klimatu. Klima čtvrtohor
✍ 551.583.4	Dendroklimatologie
✍ 551.583.7	Paleoklimatologie

ÚISK – OBA T04

9

Hierarchie a asociace v MDT - grafické vyjádření



ÚISK – OBA T04

10

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

11

11

MDT - všeobecné pomocné znaky

- = jazyka
- (0.) formy dokumentu
- () místa
- "." času
- (=) etnický
- .00 hlediska
- 03 materiálu
- 05 osob

ÚISK – OBA T04

12

12

MDT - zvláštní pomocné znaky

-1/-9 analytické

.01/.09 opakující se charakteristiky

'1/'9 syntetické

13

MDT - způsob vyjádření složených témat

1) Fasetový

⇒ rozvinuté znaky

a) kombinace hlavních a pomocných znaků

b) kombinace faset v tabulkách hlavních znaků

2) Syntaktický

⇒ složené znaky

koordinace pomocí spojovacích symbolů

14

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 | skrytí 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 mířta. Tabulka 1e
—(=..)— Všeobecné pomocné znaky plemen, etnických skupin a národností. Ta
—...— Všeobecné pomocné znaky času. Tabulka 1g
—0..— Všeobecné pomocné znaky obecných charakteristik: vlastností, materiá

Hlavní tabulky

0 VĚDA A VĚDĚNÍ. ORGANIZACE. VĚDA O POČÍTAČÍCH. VÝPOČETNÍ TECHNIKA. II
1 FILOZOFIE. PSYCHOLOGIE
2 NÁBOŽENSTVÍ. TEOLOGIE
3 SPOLEČENSKÉ VĚDY
5 MATEMATIKA. PŘÍRODNÍ VĚDY
6 APLIKOVANÉ VĚDY. LÉKAŘSTVÍ. TECHNIKA
7 UMĚNÍ. REKREACE. ZÁBAVA. SPORT
8 JAZYK. LINGVISTIKA. LITERATURA
9 GEOGRAFIE. BIOGRAFICKÉ STUDIE. DĚJINY

formální

obsahové

15

Plně fasetové třídy v MDT

2 Náboženství. Teologie

7 Umění. Rekreační. Zábava. Sport

8 Jazyk. Lingvistika. Literatura

93/94 Dějiny

...a další

16

MDT - spojovací symboly

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

ÚISK – OBA T04

17

17

Ekvivalence, hierarchie a asociace v tezuru 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	adaptaace na změnu klimatu (cs) climate system (en) zmírnutí dopadů změny klimatu (cs)
ENTRY TERMS	① 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)

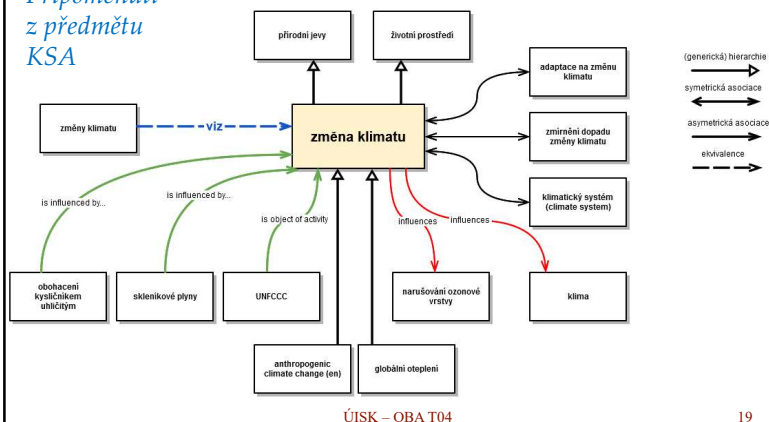
ÚISK – OBA T04

18

18

Ekvivalence, hierarchie a asociace v tezuru Agrovoc – grafické vyjádření

Připomenutí z předmětu KSA



19

19

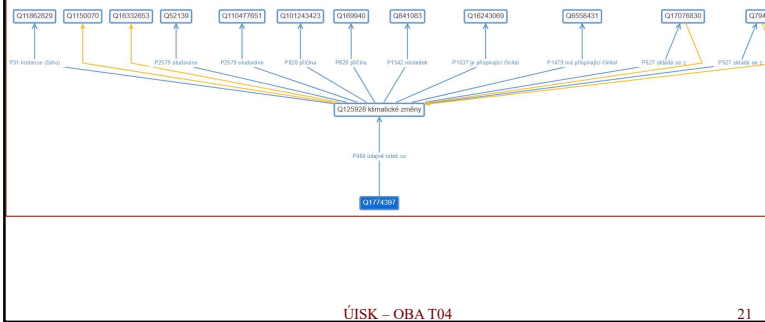
Ekvivalence, hierarchie a asociace – vyjádření ve Wikidatech

climate change (Q125928)			
change in the statistical distribution of weather patterns for an extended period, not necessarily current or recent			
climatic variation			
- In more languages			
Configure			
Language	Label	Description	Also known as
English	climate change	change in the statistical distribution of weather patterns for an extended period, not necessarily current or recent	climatic variation
Czech	klimatické změny	změny klimatu v geologické minulosti, v době historické i v současnosti na planetě Zemi	změny klimatu
Slovak	klimatické zmeny	No description defined	
German	Klimawandel	Veränderung des Klimas auf der Erde	Klimaänderung Klimawechsel Klimaveränderung
All entered languages			
Statements			
instance of	academic discipline		edit
+ 0 references			

20

20

Ekvivalence, hierarchie a asociace ve Wikidatech – grafické vyjádření



21

Systemy organizace znalostí podle vyjádření syntagmatických vztahů

JAK

- enumerativní
- analyticko-syntetické } expanzivní (otevřené)
- fasetové

KDY

- prekoordinované
- postkoordinované

ÚISK – OBA T04

22

22

MDT – spojovací symboly

Nástroje vyjádření syntagmatických vztahů (*aplikační syntax*)

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

ÚISK – OBA T04

23

23

System organizace znalostí jako model metadat

ÚISK – OBA T04

24

24

Připomenutí z předmětu KSA

Model je...

(1) Agent vytváří/používá model
 (2) Model zastupuje pro agenta originál
 (3) Agent tvoří či mění originál podle modelu
 (4) Originál je podobný modelu
 (5) Model je podobný originálu

ÚISK – OBA T04 25

25

Připomenutí z předmětu KSA

Pojmový model je... *conceptual model*

CO KDO PRO KOHO ČÍM K ČEMU
 ↓ ↓ ↓ ↓ ↓
cokoli kdokoli pro kohokoli k čemukoli

- Pojmový model je typ modelu tvořený **pojmy**.
- Pojem je nesdělitelný (implicitní).
- Pro komunikaci a manipulaci je nutné vyjádřit pojem nějakým **znakem/jazykem**.
- Systém organizace znalostí je typ pojmového modelu, který modeluje metadata (vztahující se k zaznamenaným znalostem).**

ÚISK – OBA T04 26

26

Obsahová analýza jako tvorba pojmového modelu

originál → **subjektivní pojmový model** → **graf** / **objektivní pojmový model = obsahová metadata**

ÚISK – OBA T04 27

27

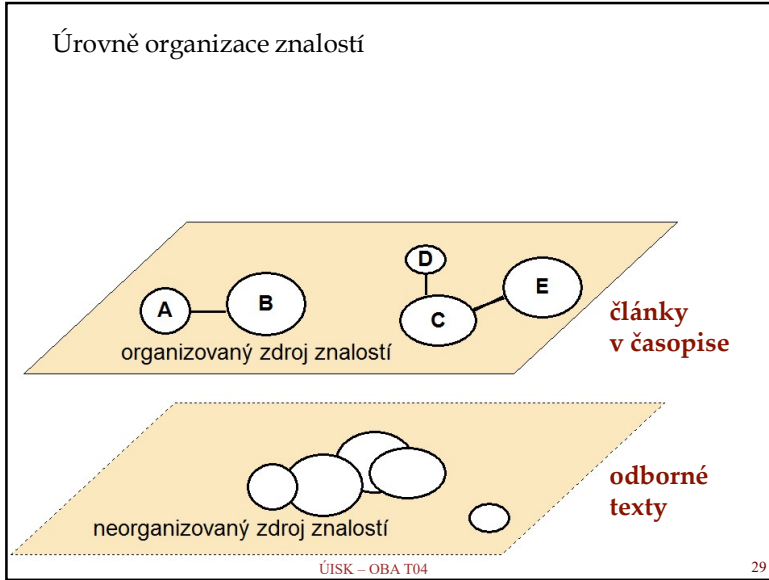
Úrovně organizace znalostí

neorganizovaný zdroj znalostí → zaznamenaná znalost (svět 3)

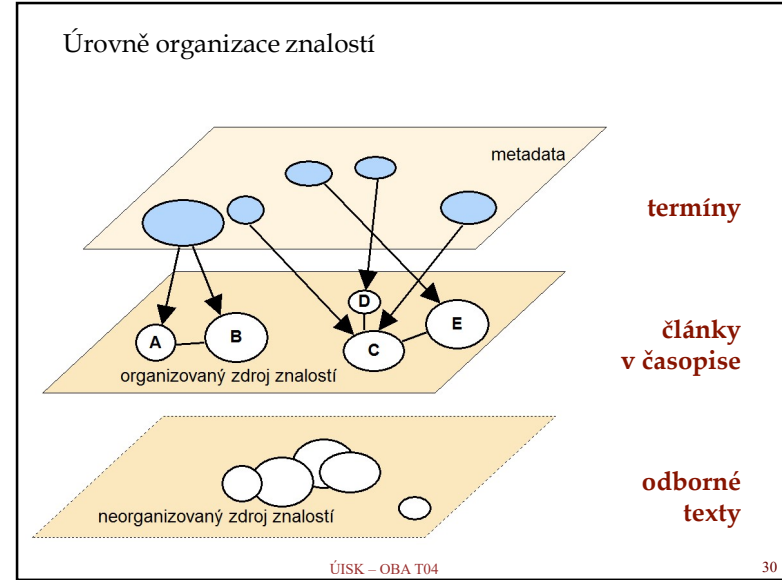
svět (svět 1) → poznatek o světě (svět 2)

ÚISK – OBA T04 28

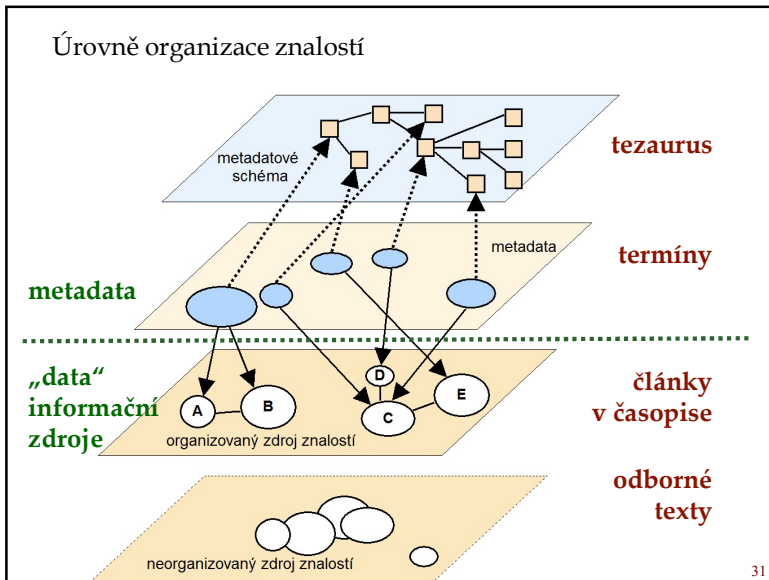
28



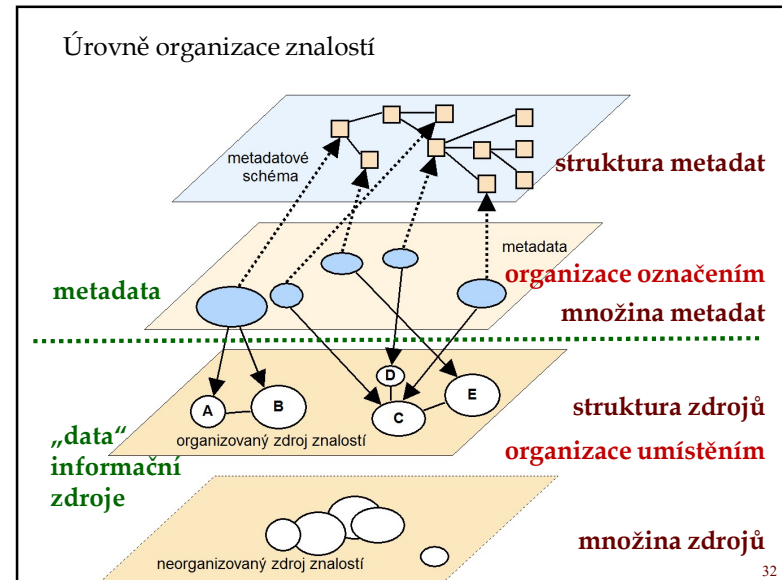
29



30



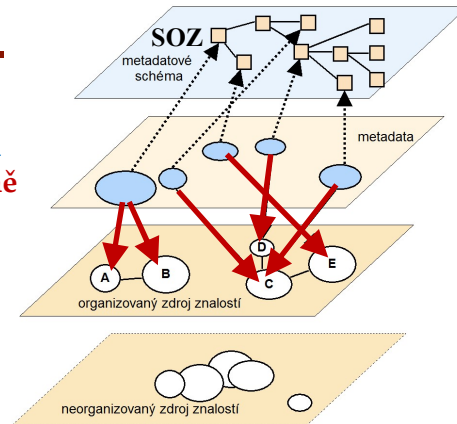
31



32

Účel systému organizace znalostí

Pojmová
indexace zdrojů
v určené doméně

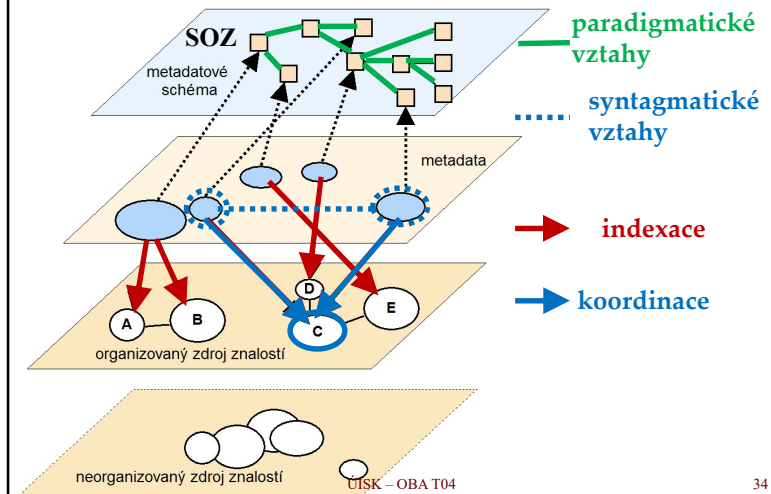


ÚISK – OBA T04

33

33

Paradigmatické a syntagmatické vztahy vyjadřované systémem organizace znalostí



ÚISK – OBA T04

34

34

Systém organizace znalostí jako řízený slovník

ÚISK – OBA T04

35

35

Systém organizace znalostí jako řízený slovník

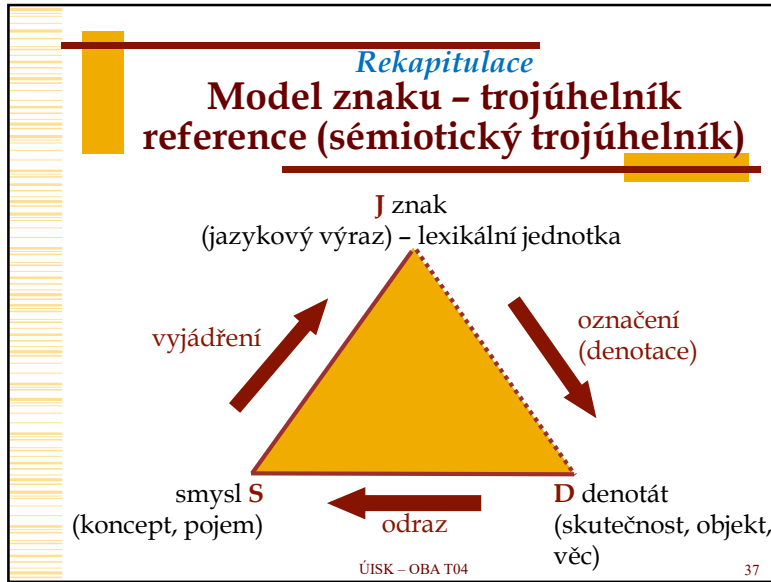
Funkce: vyjádření sémantiky i syntaxe organizovaného celku, případně i pravidel, určujících jeho používání.

- 1) **Definuje obsah** / význam zahrnutých pojmů
- 2) **Definuje rozsah** (zahrnuté / nezahrnuté pojmy)
- 3) Řeší **synonymii a homonymii** (tj. vztahy znak-pojem)
- 4) Poskytuje **kontext** potřebný pro pochopení smyslu zpracovávaných informačních objektů (tj. paradigmatické vztahy pojem-pojem)
- 5) **Definuje formát** znaků (slovosled, singulár/plurál ad.)

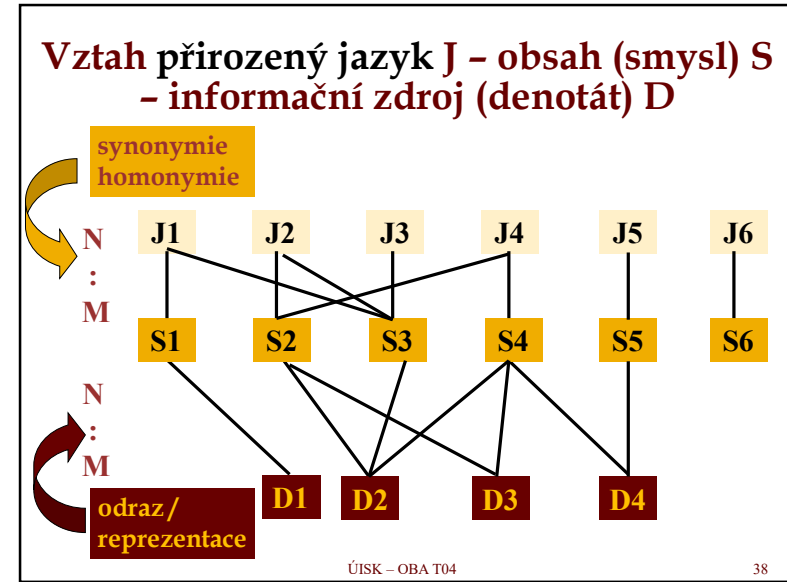
ÚISK – OBA T04

36

36



37

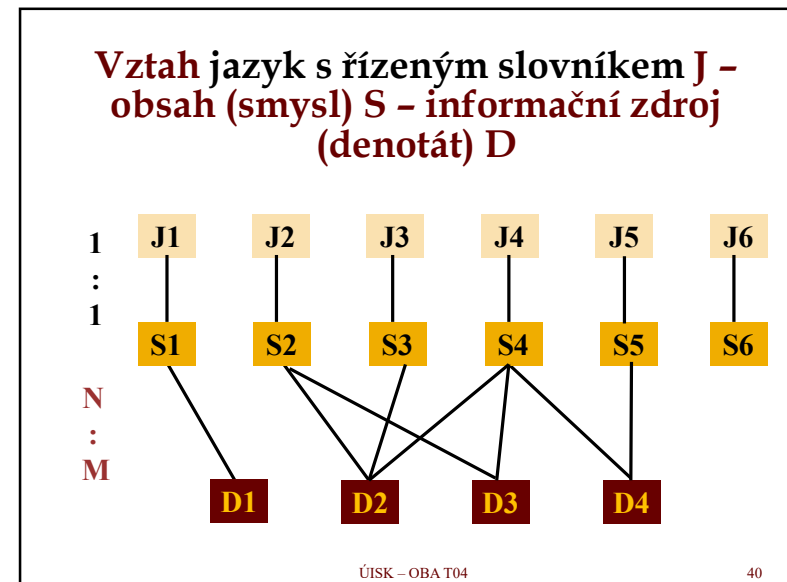


38

Název	Organizace znalosti: klíčová témata / Helena Kučerová
Kučerová, Helena, 1954- Organizace znalosti: klíčová témata / Helena Kučerová. -- První vydání. -- Praha : Univerzita Karlova, nakladatelství Karolinum, 2017. -- 269 stran : grafy, tabulky ISBN 978-80-246-3587-3	017
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: REI General; Kno	MESH deskriptor informační věda sémantika data mining management znalostí
3. Organizace znalostí Kučerová	Publikační typ monografie učebnice vysokých škol knihovnictví, informační věda a muzeologie
By: Kučerová, Hel Czech, Databáze Konspekt	NLK obory 00
Témata: dovednosti a znalosti; informace; informační management; informační zdroje; organizace informací	

ÚISK – OBA T04 39

39



40

Proč potřebujeme slovník



Disk z Faistu
1700 př. n. l.

Samotná data nestačí

10 362	10 364	10 313
5 036	5 037	5 006
5 326	5 327	5 307

ÚISK – OBA T04

41

41

Proč potřebujeme slovník



Disk z Faistu
1700 př. n. l.

Tab. 01.01 Vybrané demografické údaje
v České republice

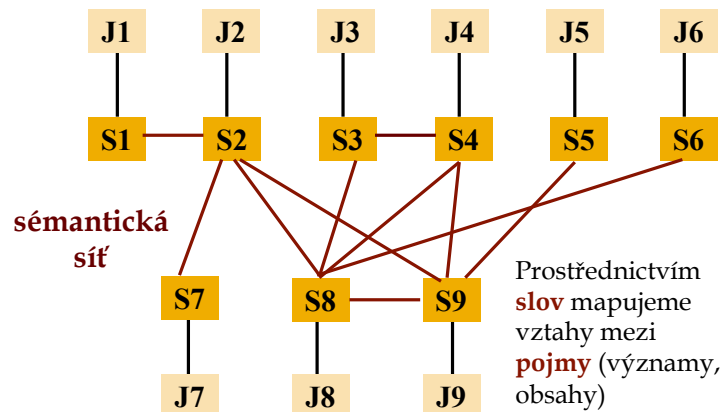
	1989	1990	1991
Počet obyvatel k 31. 12. (v tis. osob)			
muži	10 362	10 364	10 313
ženy	5 036	5 037	5 006
	5 326	5 327	5 307

ÚISK – OBA T04

42

42

Vztah řízený slovník- obsah (smysl)



ÚISK – OBA T04

43

43

Podmínka komunikace – jednotný kód (tj. slovník)



ÚISK – OBA T04

44

44

Přirozený jazyk

Systém

- **univerzální** (neomezený obsahem sdělení)
- **složitý**, jeho používání vyžaduje **inteligenci**
- **otevřený, neustále se měnící**

Vývoj přirozeného jazyka

- **smskovat, vygooglit**
- **čanit, lepá**
- **email, žába, kočka**

ÚISK – OBA T04

45

45

Problémy (nejen počítačů) s přirozeným jazykem

žena

substantivum

ženou

← →

hnát

verbum

?

?

A jedli a pili a hodovali . Přišel bratr Marek se **ženou** a dětmi , i čeládku přivedl , zadarmo se nacpat , navozené do cesty a zdálky slabé vrčení aut , která se **ženou** po státní silnici . Mžikla jsem očima dozadu po holkách svobodný , zapletený do poměru s Margitkou , vdanou **ženou** , a nevěřící už na nic , nic neberoucí příliš vážně druhé manželství trvá osm dní . Známost s mou druhou **ženou** trvala dva roky , s první šest let . 28 . září Bolševik rodiče s její volbou souhlasili , a tak se stane první **ženou** v dějinách rodiny , která se provdá z lásky . V teokratickém krejzlík pod krkem a zášť jim číší z očí , a žáčka **ženou** postrkem , jen do dveří jim vkročí , sluněčkům našim

úplně jsem bésnil a věděl , že my to vlastně zkusíme **hnát** , až kam to jde , vyzkoušet pružinu , vlastně jsme bylo jako . . . kostra Josef Novák ze země zdvih **hnát** , myslím , že to mohla být mužská holenní kost , objevil , měl jsem o něj přerazit tyč , měl jsem ho **hnát** sviňským krokem ! O dvě hodiny později jsme stále

ÚISK – OBA T04

46

46

Řešení složitosti přirozeného jazyka

- 1) **Formalizace/standardizace** jazyka/slovníku
 - speciální jazyk (terminologie)
 - řízený slovník
 - umělý jazyk

Příklad formalizace jazyka:

[Basic English](#)
[Simple English Wikipedia](#)

Simple English Wikipedia



- 2) „**Hrubá síla**“ (*brute force*)
porovnávání textových řetězců
- 3) **Počítačové zpracování přirozeného jazyka (NLP, LLM...)**

ÚISK – OBA T04

47

47

Možnosti řešení lingvistických a sémantických problémů

KDO/CO

- 1) na straně systému
- 2) na straně zpracovatele / uživatele

KDY

- 1) při tvorbě SOZ
- 2) během indexace zdroje
- 3) během požadavku na zdroj

ÚISK – OBA T04

48

48

Nástroje řešení lingvistických a sémantických problémů

JAK

1) řízený slovník (SOZ)

- zjednoznačnění významu
 - vyjádření vztahů
- } sjednocení sémantiky

2) indexový soubor

- z uložených dokumentů
 - z realizovaných dotazů (seznam uložených dotazů)
- } přístup

3) kontext

- kolokabilita – v textu, v indexu
- převod syntagmatických vztahů na paradigmatické

ÚISK – OBA T04

49

49

Nástroje rozšiřování požadavků pro jednotlivé typy synonymie - 1

1) lexikální synonymie

řízený slovník

enumerace: předem vytvořené seznamy synonym s definovanými paradigmatickými vztahy (tezaury, authority, {synset})

jednotné označení: preferované termíny, notace, identifikátory

⇒ rozšiřování (expanze) dotazu o synonyma:

horizontální (ekvivalence)

vertikální (hierarchie)

asociativní

Příklady:

termín „knihovnictví“ v EUROVOCu

třída „knihovnictví“ v MDT

ÚISK – OBA T04

50

50

Nástroje rozšiřování požadavků pro jednotlivé typy synonymie - 2

2) gramatická synonymie

řízený slovník

enumerace: předem vytvořené seznamy slovních kmenů a tvarů + systém pravidel odvozování

1. automatický převod na základní tvar (lemmatizace)
2. odvození možných tvarů (derivace)

⇒ rozšiřování (expanze) dotazu o další tvary slova

Příklad: sloveso „jít“ v Českém národním korpusu

ÚISK – OBA T04

51

51

Nástroje rozšiřování požadavků pro jednotlivé typy synonymie - 3

3) „nouzové řešení“, formální synonymie

indexový soubor

zkracování (a rozšiřování) koncovek / maskování

⇒ automatické doplňování, našeptávač / *autocomplete*, oprava překlepů

ÚISK – OBA T04

52

52

Problém nástrojů formálního rozšiřování dotazu (*brute force*)

The screenshot shows the ProQuest search interface. At the top, there are navigation links: 'Jednoduché hledání', 'Pokročilé hledání', 'Publikace', 'Procházet', and 'Data báze (2)'. The search bar contains the text 'FRBR IFLA LRM'. Below the search bar, there are two checkboxes: 'Plný text' (checked) and 'Recenzované' (unchecked). A yellow suggestion box below the search bar displays the text 'Měli jste na mysli: froe fula wrm'.

ÚISK – OBA T04

53

53

Nástroje omezování požadavků při řešení homonymie

1. řízený slovník

- a) homonyma vyloučíme z používání (nahrazení jednoznačným synonymem)
balón viz míč
- b) jednotlivé významy upřesníme poznámkami (kvalifikátory, alternativní definice)
balón (míč)

2. kontext (= syntagmatické ad hoc vztahy)

balón v Českém národním korpusu

ÚISK – OBA T04

54

54