

## Téma 2 – Lingvistické a sémantické problémy obsahové analýzy

1. Co je to obsahová analýza (OA)?
2. **Ontologie OA**  
Co je to obsah?  
Kde je obsah?
3. **Gnozeologie OA**  
Jak zjistit obsah?  
Jak vyjádřit obsah?
4. **Praxe OA**  
Jak chápat/modelovat obsah?

ÚISK – OBA T02

1

1

## Gnozeologické problémy obsahové analýzy

- **Kognitivní** – jak to poznat?
  - Jak zjistit potřebu informace (= obsah informační potřeby)
  - Jak zjistit obsah informačního zdroje
- **Sémanticko-lingvistické** – jak to vyjádřit?
  - Jak vyjádřit potřebu informace
  - Jak vyjádřit obsah zdroje

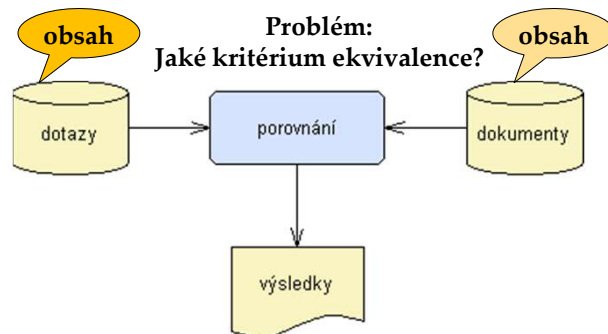
ÚISK – OBA T02

2

2

## Problémy s obsahem

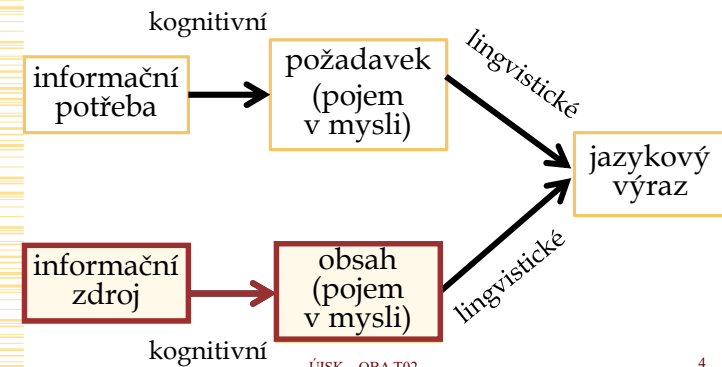
### Obecný model vyhledávacího systému



3

3

## Gnozeologicko-sémantický šum - komponenty a transformační procesy



ÚISK – OBA T02

4

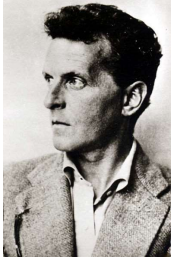
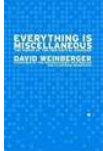
4

## Lze poznat obsah / smysl / význam / předmět? 3

letní sleva na **oblečení**      1) Význam určuje kontext  
sleva na letní **oblečení**

2) Význam se určuje používáním  
Ludwig Wittgenstein (1889–1951)  
„Meaning is determined by use“

3) David Weinberger  
Everything is miscellaneous. 2007  
<http://www.everythingismiscellaneous.com/>

ÚISK – OBA T02 5

5

## Jak určit / zjistit obsah informačního zdroje? 3



Jens-Erik Mai, 1968  
profesor informačních studií  
na Univerzitě v Kodani

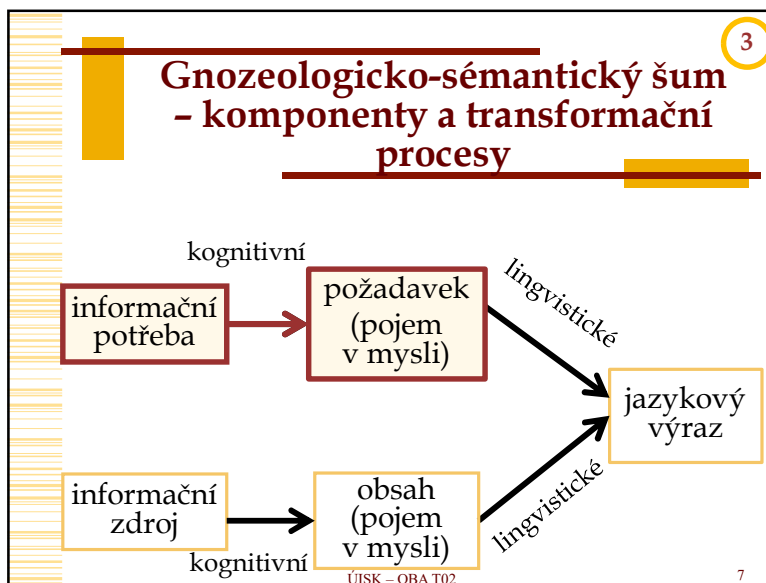
**Co je to?**

„Předmět neexistuje v dokumentu, připraven k objevení a vyzvednutí indexátorem. Indexátor bude nezbytně vkládat do své analýzy dokumentu znalosti nebo odhady o potenciálním užití dokumentu a uživatelských potřebách.“

MAI, Jens-Erik. Analysis in indexing: document and domain centered approaches. In: *Information processing and management*. May 2005, 41(3), 603. ISSN 0306-4573.

ÚISK – OBA T02 6

6



7

## Jak určit / zjistit potřebu informace? 3

„A jakým způsobem budeš hledat, Sókrate, to, o čem vůbec nevíš, co to je? Kterou věc z těch, které neznáš, uděláš předmětem svého hledání? Nebo i kdybys nakrásně na ni připadl, jak budeš vědět, že to je to, co jsi neznal?“

Platón: Menón

↓

- 1) Na **neznámé** se ptáme prostřednictvím **známého**, na „**nové**“ se ptáme prostřednictvím „**starého**“
- 2) Nesnažíme se dosáhnout totožnosti, ale **podobnosti** dotazu a vyhledané informace

ÚISK – OBA T02 8

8

### Jak určit / zjistit potřebu informace? 3

Jak získá směr, popíše zakázku

Jak to vedoucí zakázky pochopí

Jak to projektce navrhne

Jak to v dílně vyrobí

Jak to obchodní zástupce popíše

Jak je zakázka zadána umělcem

Co je vše do dílny

Jak je získá směr vyfakturován

Jaká je podpora zakázky

Co získá směr vlastně potřeboval

9

9

### Jak určit / zjistit potřebu informace? 3

**serendipity**

„Je to často položka, která **neodpovídá** našemu zájmovému rešeršnímu profilu, jež se ukáže jako nejzajímavější; mnohé z nejvýznamnějších vědeckých objevů vznikly jako výsledek šťastné náhody ... a systém, který vylučuje tuto možnost, může být až příliš úspěšný v tom, jak odpovídá potřebám vyjádřeným uživatelem!“

FOSKETT, Anthony Charles. *The subject approach to information*. 5th ed. London: Library Association Publishing, 1996, s. 26.

ÚISK – OBA T02 10

10

### Gnozeologicko-sémantický šum - komponenty a transformační procesy 3

informační potřeba → požadavek (pojem v myslí)

informační zdroj → obsah (pojem v myslí)

požadavek (pojem v myslí) → jazykový výraz (lingvistické)

obsah (pojem v myslí) → jazykový výraz (lingvistické)

⇒ Cvičení 9

ÚISK – OBA T02 11

11

teorie poznání  
gnozeologie  
epistemologie  
kognitivní věda  
noetika

komunikace  
sociologie  
etnologie

sémiotika

sémantika

lingvistika

gnósis  
γνώσις

epistémé  
ἐπιστήμη

noéma  
νοήμα

pragmatika

fonetika  
grafémika

lexikologie

pravopis

morfologie

syntax

matematická  
lingvistika

počítačová  
lingvistika

kognitivní  
lingvistika

jazykové plány / roviny

ÚISK – OBA T02 12

12

## Systémové pojetí jazyka

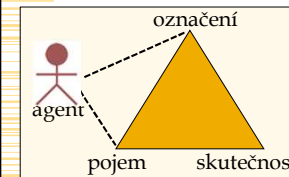
prvky	lexikální subsystém ( <u>slovník</u> )	
	pravidla (gramatika, syntaxe)	
vztahy	paradigmatické	
	syntagmatické	
funkce	poznávací	referenční
	komunikační	
	jazyk = skutečnost myšlenky	
	emotivní, poetická, konativní ad.	

ÚISK – OBA T02

13

13

## Typologie jazykových systémů



- čestina
- TDKIV
- AGROVOC
- HTML

- Přirozený jazyk**  
„měkký“, fuzzy systém
- Speciální jazyk (terminologie)**
- Jazyk s řízeným / strukturovaným slovníkem**  
jednoznačné prvky
- Umělý jazyk**  
„tvrdý“ systém  
platí pravidla formální logiky  
jednoznačné prvky i vztahy  
notace

ÚISK – OBA T02

14

14

## Typologie vztahů v jazyce

<b>Paradigmatické</b> sémantické trvalé, <i>a priori</i> bezkontextové	<b>Syntagmatické</b> syntaktické ad hoc, <i>a posteriori</i> kontextové
organizace označením vztahy <b>významů</b> slov např. ve slovníku	organizace umístěním vztahy <b>umístění</b> slov v promluvě / v textu
<i>langue</i> (systém jazyka)	<i>parole</i> (promluva, mluva)
<ul style="list-style-type: none"> <li>ekvivalence</li> <li>hierarchie (obecná / generická, partitivní, instanční)</li> <li>asociace</li> </ul>	<ul style="list-style-type: none"> <li>kolokace</li> <li>koordinace</li> </ul>

ÚISK – OBA T02

15

15

## Paradigmatické a syntagmatické vztahy

skákal	pes	přes oves	osa substituce (paradigmatická)  význam nezávislý na kontextu
poskakoval	Alík	ovesným polem	
běžela	šelma	skrz pole	
	psovítá	s obilím	
osa kombinace (syntagmatická)			

ÚISK – OBA T02

16

16

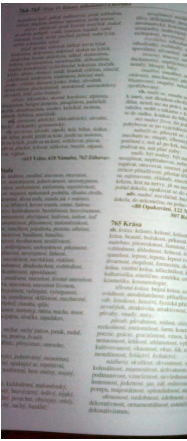
## Paradigmatické a syntagmatické vztahy

skákal	pes	přes oves	<b>osa substituce (paradigmatická)</b> <i>význam nezávislý na kontextu</i>
poskakoval	Alík	ovesným polem	
běžela	šelma psovitá	skrz pole s obilím	
život je	pes	psí počasí	
<b>osa kombinace (syntagmatická)</b> <i>význam určený kontextem</i>			

ÚISK – OBA T02 17

17

## „Pěkný“ v Tezauru jazyka českého paradigmatické vztahy (a priori)



**krásný**, překrásný, čarokrásný, kouzelný, **pěkný**, šumný, onaký, hezký, přehezský, hezounký, hezoučký, báječný, fantastický, znamenitý; **půvabný**, okouzlující, milý, roztomilý, slíčný, lepý, ladný, spanilý, líbezný, rozkošný, činaný, hezký jako obrázek, malovaný, k pomilování, šarmantní, gráciézní; **pohledný**, vzhledný, úhledný, švarný, fešný, dobře vypadající, vznosný, urostlý, ztepilý, souměrný, útvárný, pěkně tvarovaný, dobře rostlý, dobře stavěný, udělaný, formózní; **přitažlivý**, zaoblený, vnaďný, bujný, bujných tvarů, lepých tvarů, sošný, svůdný, rajcovný, ženský, sexy, atraktivní, beroucí dech; **elegantní**, vytříbený, vybraný, dokonalý, vkusný, módní, šik, apartní, mondénní, nenucený, uhlazený, dobře udržovaný, pěstěný, upravený, úhledný, jako ze škatulky, vystrojený, vyšňořený, švihácký, fešácký; **nádherný**, velkolepý, znamenitý, skvělý, báječný, kultivovaný, oslnivý, zářivý, úchvatný, senzační, úžasný, božský, kouzelný, skvostný, splendidní, magnificentní;

ÚISK – OBA T02 18

18

## Paradigmatické vztahy

Připomenutí z předmětu KSA

<b>ekvivalence</b>	obsahu
	formy
<b>hierarchie</b>	<b>obecná</b> (širší – užší extenze / intenze)
	<b>závislost</b> (když..., tak...), <b>historická</b> (předchozí – následující)
	<b>„speciální“</b> (s inferencí a dědičností) <ul style="list-style-type: none"> <li>■ generická (rod – druh)</li> <li>■ partitivní (celek – část)</li> <li>■ instanční (obecné – individuální)</li> </ul>
<b>asociace</b>	sekvenční (prostor, čas)
	příčinná
	...

ÚISK – OBA T02 19

19

## Ekvivalence v synonymii

J1

J2

S

D

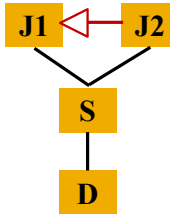
2406 **ZNAK** (logický) čeho (pojmu, druhu, jedince, pravdy ap.; z. základní, podstatný, nezbytný, nutný, odlišující, roz-, význačný, charakteristický, jistý, bezpečný, nemýlný, neo-), podstatná, příznačná, charakteristická **známka**, znamení čeho (vzdělanosti, poctivosti), **význačná**, rozlišující **vlastnost**, **podstatný**, charakteristický **rys** čeho, *kniž.* význak, **přívlastek** (p-y lidské; p-y hluboké povahy), **atribut** (pojmu, realismu), **charakteristikon** (narodnosti, génia), **signum** (neomylné s. pravého umění), **pečet** (individuality, pravdy), *zast.* **oznak** (hlavní o. pravdy), **přímět(ek)**, *stř.* **poznatek** (egoismu); *log. z.*, určující prvek při stanovení pojmu  
*viz též* typ, typický

ÚISK – OBA T02 20

20

## Hierarchie v synonymii

archilexém



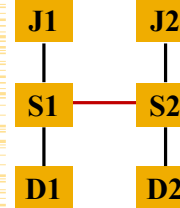
2406 **ZNAK** (logický) čeho (pojmu, druhu, jedince, pravdy ap.; z. základní, podstatný, nezbytný, nutný, odlišující, roz-, význačný, charakteristický, jistý, bezpečný, nemylný, neo-), podstatná, příznačná, charakteristická **známka**, znamení čeho (vzdělanosti, poctivosti), **význačná, rozlišující vlastnost, podstatný, charakteristický rys** čeho, *kniž.* **význam, přívlastek** (p-y lidské; p-y hluboké povahy), **atribut** (pojmu, realismu), **charakteristikon** (národnosti, génia), **signum** (neomylné s. pravého umění), **pečet** (individuality, pravdy), **zastl. oznak** (hlavní o. pravdy), **přímět(ek)**, *zř.* **poznatek** (egoismu); *log. z.*, určující prvek při stanovení pojmu  
*viz též typ, typický*

ÚISK – OBA T02

21

21

## Asociace v synonymii



2406 **ZNAK** (logický) čeho (pojmu, druhu, jedince, pravdy ap.; z. základní, podstatný, nezbytný, nutný, odlišující, roz-, význačný, charakteristický, jistý, bezpečný, nemylný, neo-), podstatná, příznačná, charakteristická **známka**, znamení čeho (vzdělanosti, poctivosti), **význačná, rozlišující vlastnost, podstatný, charakteristický rys** čeho, *kniž.* **význam, přívlastek** (p-y lidské; p-y hluboké povahy), **atribut** (pojmu, realismu), **charakteristikon** (národnosti, génia), **signum** (neomylné s. pravého umění), **pečet** (individuality, pravdy), **zastl. oznak** (hlavní o. pravdy), **přímět(ek)**, *zř.* **poznatek** (egoismu); *log. z.*, určující prvek při stanovení pojmu  
*viz též typ, typický*

ÚISK – OBA T02

22

22

## „Pěkný“ v Českém národním korpusu syntagmatické vztahy (a posteriori)

Shuffle: ✓

whole "omezeni/syn2015") | ARF: 1,076.41 | Result is shuffled

1 / 55

, ale spíš sháním něco jiného . Když uvidím nějaký **pěkný** kousek za rozumnou cenu a mám peníze , rá  
 . Jestli o nás jenom cekne , tak poletí z **pěkný** vejšky . "<p><p>„ Jo , ale ...  
 ( o něm více v následující kapitole ) , upoutá **pěkný** starý dům U Splavninů čp. 36/III – s podlouhým  
 hodinky přímo od dealerů . Ti , když mají nějaký **pěkný** model , se na mě už rovnou obrací . "<p>  
 jaký lidi potkává na chodbě . " Aspoň jsou tam **pěkný** holky , " řekl jsem jí , když odcházela .  
 jakousi bedničku a pronesl improvizované uvítání . Byl to velice **pěkný** , spontánní úvod k samotnému zahájení . Dř  
 Meil stála u otcovy ledové rakve .<p><p>Byl na ni **pěkný** pohled , ale Žestokost se jen s pohledem ne  
 si neřekla : „ Tak a teď si udělám svůj **pěkný** portrét a dám ho na odív . " Spíše na  
 to rohový pozemek a má kouzelnou zahrádku . Květiny a **pěkný** trávník . Chtějí za něj tři sta tisíc . Ale  
 která právě unikla smrti . K nejbližšímu záhybu chodby doběhla **pěkný** kus před Valerí , a když k němu dorazila vzle  
 , který skončil v síti . Druhý už byl ale **pěkný** . Doufám , že tak budu pokračovat i dál .  
 ÚISK – OBA T02

23

## Jak vyjádřit obsah / potřebu informace?

3

Lingvistické a sémantické problémy organizace znalostí – asymetrické vztahy 1 : N

- 1) **synonymie** (1 význam – N lexémů)
- 2) **homonymie** (1 lexém – N významů)  
XOR, buď / nebo, vylučující
- 3) **koordinace – složené pojmy / složená témata** (N významů)  
AND, a zároveň, partitivní
- 4) **slovní spojení** (1 význam – N lexémů)

ÚISK – OBA T02

24

24

## Co je to lexikální jednotka? též **lexém**, *lexical unit, token*

### obecně

- sekvence písmen, číslic a dalších specifických znaků používaná pro označení určitého **pojmu**
- jednotka jazyka, která je nositelem významu nebo funkce
- z hlediska významového tvoří jeden nedělitelný celek

### v organizaci znalostí

- slovo nebo slovní spojení ve formě buď
  - výrazu přirozeného jazyka, nejčastěji substantiva nebo substantivního spojení
  - notačního znaku

ÚISK – OBA T02

25

25

## Lexikální subsystém jazyka – způsoby zaznamenání

**Korpus**  
(*parole*)

**Tezaurus**  
(*langue*)

vztahy

syntagmatické

paradigmatické

„přirozený jazyk  
v počítači“  
– Český národní korpus  
<http://www.korpus.cz/>



ÚISK – OBA T02

26

26

## Výkladový slovník vs. tezaurus

slovník	tezaurus
způsob použití	
slovo (znak)	pojem (význam)
Co znamená? ↓	↓ Jak vyjádřit?
pojem (význam)	slovo (znak)
základní (primární) uspořádání	
abecední	systematické
základní (primární) proces	
sémaziologický	onomaziologický

ÚISK – OBA T02

27

27

## Peter Mark Roget 1779 – 1869



Anglický polyhistor:

lékař, lexikograf, autor četných publikací z oboru medicíny a optiky, přispěvatel do encyklopedií, vynálezce logaritmického pravítka...

*Thesaurus of English Words and Phrases, Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary Composition.*

1. vydání 1852

- slovník synonym a antonym
- kategorizace slov

ÚISK – OBA T02

28

28

## Roget's Thesaurus

**Main Entry:** cat  
**Part of Speech:** noun  
**Definition:** feline animal, sometimes a pet  
**Synonyms:** bobcat, cheetah, cougar, grimalkin, jaguar, kitten, kitty, leopard, lion, lynx, malkin, mouser, ocelot, panther, puma, puss, pussy, tabby, tiger, tomcat

Zdroj: Roget's New Millennium™ Thesaurus, First Edition (v 1.3.1)  
 Copyright © 2006 by Lexico Publishing Group, LLC. All rights reserved.

ÚISK – OBA T02 29

29

## Roget's Thesaurus

*Existence*

*One, entity, being, exist<sup>ce</sup>, essence, quintess<sup>ce</sup>, quidd<sup>ce</sup>, nature, thing, substance, course, world, frame, reality, (v. truth) actual exist - fact, course of things, under, extant, present*

*Positive, affirmative, absolute, intrinsic, substantive, to be, exist, obtain, stand, pass, subsist, prevail, lie - on fact, on y tapis, to constitute, form, compose, state, hold, exist<sup>ce</sup>, condition, predicament, situation, posture, circumstances, case, plight, term, tune, -point, degree, juncture, conjuncture, pass, emergency, exigency*

Existence

End, entity, being, exist <sup>ce</sup>	Nonentity, nullity, nihility
Essence, quintess <sup>ce</sup> , quidd <sup>ce</sup>	nonexist <sup>ce</sup> , noth <sup>g</sup> , nought
Nature, thing, substance, course, world, frame, reality, (v. truth) actual, exist <sup>ce</sup> - fact, course of things, under y sure extant, present	void, zero, cypher, blank, empty, unsubstantial, imaginary, visionary, fabulous, fictitious, supposititious, absent, shadow, dream, phantom, phantasm, Negation, virtual, extrinsic, potential, adjective
Positive, affirmation, absolute, intrinsic, substantive, to be, exist, obtain, stand, pass, subsist, prevail, lie - on fact, on y tapis, to constitute, form, compose, to consist of	inherent, obtain, stand, prevail, lie - on fact, on y tapis, scope, habits, temperament, nature, constitut <sup>n</sup> , habit, place, situation, posture, contingency
Slate, mode of exist <sup>ce</sup> , condition, predicament, situation, posture, circumstances, case, plight, term, tune, -point, degree, juncture, conjuncture, pass, emergency, exigency	

ÚISK – OBA T02 30

30

4

## Rekapitulace Model znaku - trojúhelník reference (sémiotický trojúhelník)

**J** znak  
(jazykový výraz) - lexikální jednotka

vyjádření ↗

označení (denotace) ↘

odraz ↙

**S** smysl  
(koncept, pojem)

**D** denotát  
(skutečnost, objekt, věc)

ÚISK – OBA T02 31

31

## Ideál - vzájemně jednoznačná korespondence znak - smysl - denotát

**Problém:**  
víceznačnost,  
synonymie,  
neurčitost  
významu

kyselina sírová	pondělí
J1	J2
S1	S2
D1	D2

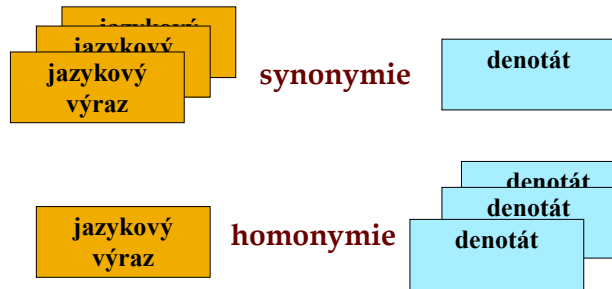
ÚISK – OBA T02 32

32



## Problém – nejednoznačnost, neurčitost významu

### Typy ekvivalence v jazyce



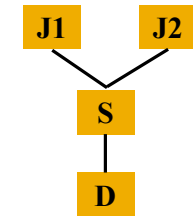
ÚISK – OBA T02

33

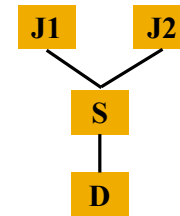
33

## Synonymie

fotbal – kopaná



doktor – lékař



ÚISK – OBA T02

34

34

## Typy synonymie

- lexikální (věcná, sémantická)**
  - „pravá“ synonymie (ekvivalence)  
*cesta – silnice – dálnice – vozovka – stezka – pěšina*
  - hierarchie  
*domácí zvíře – kočka – Macek*
  - asociace  
*nevolnost – nemoc, plný – prázdný*
- mezijazyková lexikální ekvivalence**  
*kočka – die Katze – кошка – cat*
- gramatická (morfologická)**  
1 slovo ve více tvarech – *cesta, cesty, cestě, cestu...*
- formální**  
*budova, budou, ubudou, budoucnost, vybudovat...*

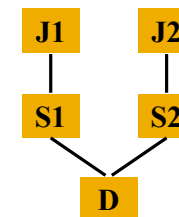
ÚISK – OBA T02

35

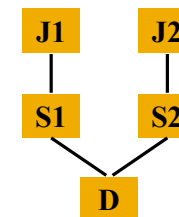
35

## Kvazisynonymie, antonymie

fotbalista – Jan Kovář



tvrdý – měkký



ÚISK – OBA T02

36

36

## Typy synonymie v bázi národních věcných autorit

### Temat. autority - Úplné zobrazení záznamu

Zvolte formát: Standardní -- MARC

Záznam 1 z 1

Ident. číslo	ph127701
Záhlaví	zvukové dokumenty
Odkaz. forma	dokumenty zvukové
Nadřazený termín	dokumenty
Podřazený termín	audioknihy
Asoc. termín	audiovizuální dokumenty katalogizace zvukových dokumentů
MDT	025.177 (knihovní fondy) 371.683 (učební pomůcky) 002.1-028.17 (obecně)
Angl. ekvivalent	sound recordings
Skupina Konspektu	025 - Organizace knihoven a knihovních fondů [12] 371 - Školství (organizace) [22]

37

37

## Kontextová sémantika - hypotézy

- **ekvivalence označením**  
slova, která se podobně píšou, mají podobný význam
- **ekvivalence umístěním**  
slova, která jsou v textech blízko sebe, mají významovou souvislost

### škola (W) hrou

Až tento školní rok se pro Petra stala **škola hrou** a poprvé přinese domů samé jedničky. V |  
vzpomene si i na třídního Ritu a jeho **školu hrou**. Druhý den je na vratech gymnázia seznam  
národ Komenského, pro kterého byla **škola hrou**. Já tu hru na školách moc často nevidím.  
stně jsem školu neměl rád! Nebyla to **škola hrou**, jako jí není ani teď. Už samotný výhled  
vi Komenskému a jeho scholé ludus, **škole hrou**. Učení je prostě efektivnější, pokud se je  
kjmí jsou trojivá škola, činná škola, **škola hrou**, otevřená, volná, svobodná, netradiční ál  
mi řemesly a pracovišti v duchu hesla **škola hrou**. Žáci takto získají konkrétní představu o u  
blesky u památníku Prokopa Diviše, **škola hrou**. \* ,propagoval učitel národů Jan Ámos Kor

### Hypotéz\*

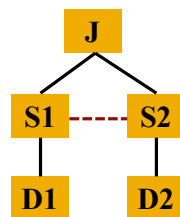
hypotéz  
Hypotéza  
hypotézách  
Hypotézám  
hypotézami  
hypotéze  
hypotézo  
hypotézou  
hypotézu  
hypotézy

38

38

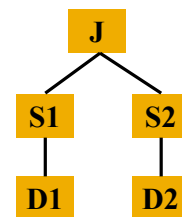
## Homonymie / mnohovýznamovost (polysémie)

jazyk - jazyk



polysémie

los - los



homonymie

ÚISK - OBA T02

39

39

## Typy homonymie

### a) lexikální

*stát - stát - stát - stát, bez - bez, Let let, léta letí, a - a*

- zvuková (*mít - mýt, ve při - vepři, sním - s ním*)
- grafická (*email - email, . - . - .*)

### b) syntaktická

*vhazování domácích*

*církev chce vrátit ukradený majetek*

*moc slov - moc slov*

### c) sémantická - ironie, metafora, metonymie

*to se ti povedlo*

*šli na skleničku*

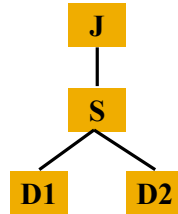
ÚISK - OBA T02

40

40

## Homonymie - mnohoznačnost (dvojsmysl)

to je pěkné - to je pěkné



ÚISK - OBA T02

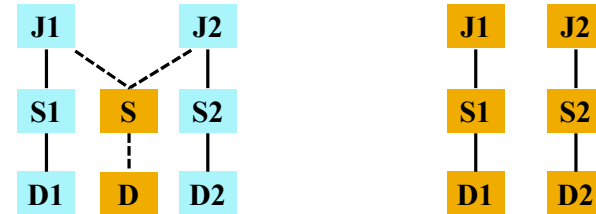
41

41

## Falešní přátelé, zrádná slova Homonyma, která se "tváří" jako synonyma

host (česky) host (anglicky)

věcné pořadí věcné pořadí



ÚISK - OBA T02

42

42

## Homonymie - metafora



Nedělala bych si marné naděje. Vítr v peněžence  
ještě nikdy žádnou vrtuli neroztočil...

43

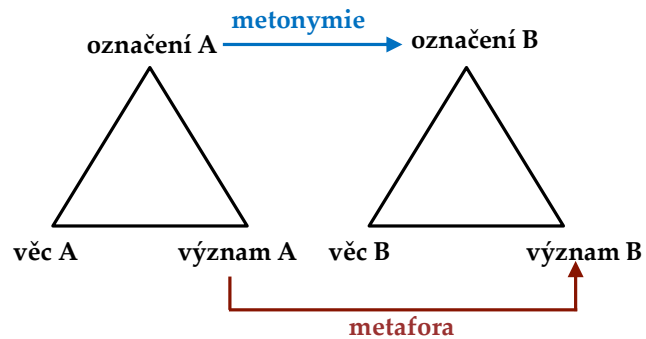
## Homonymie - metonymie

Čínská filosofie  
je v C244  
(1. patro vpravo a zadním  
schodištěm do 2.p.)

44

44

## Metafora a metonymie = přenos významu / označení

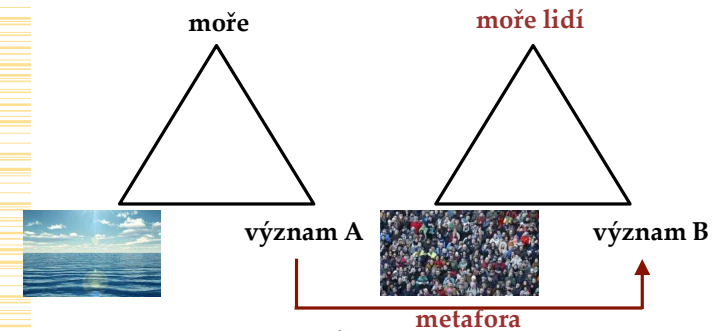


ÚISK – OBA T02

45

45

## Metafora = přenos významu

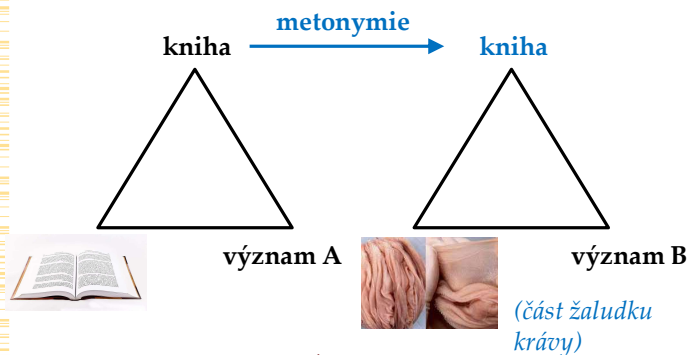


ÚISK – OBA T02

46

46

## Metonymie = přenos označení



ÚISK – OBA T02

47

47

## Složené téma / pojem

- elektrokolo
  - husopaska
  - teploměr
  - e-learning
  - samoobsluha
  - narkobaron, uhlobaron
  - zahraničněpolitický
  - alfanumerická notace
  - aktivní meziknihovní výpůjční služba
  - automatická tvorba tezauru
  - daň z přidané hodnoty
  - matematicko-statistická teorie informace
- kyber-egyptologie

ÚISK – OBA T02

48

48

## Jak poznat složené téma / pojem

### Zaměnitelnost

elektronická	kniha	příslloví
modlitební	cigareta	pro mládež
dětská	jízdenka	se slevou

ÚISK – OBA T02

49

49

## Slovní spojení

těž sousloví, sdružené pojmenování, fráze, frazém, idiom (*phrase*), n-gram, víceslovný lexém, kolokace, koherentní víceslovný výraz

*Jan Neruda, černí baroni, stavební úřad, pražské jaro, láry fáry, New York City, postavení mimo hru, lze si do vlastní kapsy*

- Ustálené pojmenování ze dvou nebo více slov, které **označuje jednu skutečnost** a vstupuje do věty jako hotový celek
- **Nejčastější forma:**  
**substantivní spojení:** lexikální jednotka, kterou lze rozčlenit na oddělené složky, přičemž každá z těchto složek může být vyjádřena substantivem a nezahrnuje sloveso ve tvaru určitém

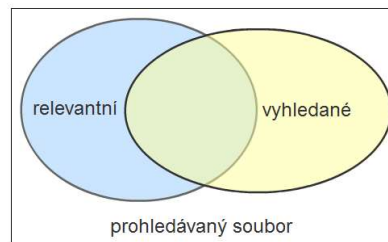
ÚISK – OBA T02

50

50

## Hodnocení výsledků vyhledávání (= výběru, prozkoumání zdrojů)

- relevance
- informační šum
- **úplnost** (*recall*)  
podíl vyhledaných z relevantních
- **přesnost** (*precision*)  
podíl relevantních z vyhledaných



**ideál:**

- najít vše, co je relevantní
- vyloučit vše, co není relevantní

ÚISK – OBA T02

51

51

## Problém synonymie

**snižuje úplnost** (*recall*):

hrozí, že nenajdeme všechny relevantní informace

**řešení:**

**rozšiřování** (expanze) požadavku o zdroje obsahující **synonyma**

ÚISK – OBA T02

52

52

## Problém homonymie

**snižuje přesnost (*precision*):**

hrozí, že najdeme nerelevantní informace

**řešení:**

**zúžení** požadavku o zdroje, obsahující nerelevantní homonyma

53

## Problém složených témat

- **analýza**
  - **syntéza**
  - **konstrukce indexu/dotazu**
  - **koordinace**
- } Jak zjistit jejich obsah / smysl?  
 Jak je vyjádřit?

54

## Problém slovních spojení

Jak poznat, že jde o slovní spojení?

Kontext

stal kohoutem . Na místě , kde předtím stál , ržál **bílý kůň** . Ted už mohla začít slavnost . Trubači zatrou  
 lo filozof a tůň Dvakrát do téže řeky Vstupuje tiše **bílý kůň** Dvakrát do řeky co je stejná Vstupuje bělouš  
 světnice , honem z tebe uteku . Snih Přišel k nám **bílý kůň** , zalehl nám celý dvůr . Snih Přiletěl k nám bí  
 kůň je vlastně pokračováním souboru Jedenáctý **bílý kůň** , který vyšel ve dvou vydáních v polovině šed  
 stníků bůhzdarma . Jenže . Nad matičkou Prahou **bílý kůň** roztáhl svá křídla pouze jedenkrát . Od této ex  
 dnikatel , který v případě nejprve vystupoval jako **bílý kůň** , vinu vzal později plně na sebe , totiž nyní za  
 míněný Václav Franta , podle odborníků typický " **bílý kůň** " , tedy nastrčená figura , která si zřejmě ani n  
 muto činu bylo pravděpodobně to , že Tulej , jako **bílý kůň** používaný při různých akcích , přestal vyhov  
 ě jeho plochého , táhlého horizontu běžel pomalu **bílý kůň** . I dívka na něm byla bílá , i vlasy vlající ve vě  
 ké fejetony , shrnuté do dvou svazků ( Jedenáctý **bílý kůň** , Třináctý černý kůň ) , zůstávají tu však jako

55