

Organizace znalostí X.

11. 12. 2020

PhDr. Jiří Stodola, PhD.

Jazyk

- Původ jazyka, teorie
 - Interjekční – výkřiky emocionálního charakteru
 - Omonatopoická – napodobování hlasů zvířat
 - Synergastická – výkřik doprovázející kolektivní akce
 - Posunková – z gest
- Monogeneze
- Polygeneze

Vlastnosti jazykového systému

- Znak musí být smysly vnímatelný a tedy sdělitelný v procesu komunikace.
- Znak je vzhledem ke svému designátu arbitrérní, je na něm nezávislý, jen na subjektu závisí, jaký znak zvolí pro označení nějaké entity.
- Znak je závazný vzhledem ke společenství, které daný znakový systém užívá, jinak by vněm nebyla možná komunikace.
- Znak je závazný vzhledem k systému, do kterého náleží.

Osy jazyka

- Paradigmatická – jazyk jako systém, osa výběru
- Syntagmatická – promluva, osa kombinace

Druhy jazyka

- Přirozený – „jazyk, který se vyvíjí a jehož pravidla vyplývají z úzu, takže nemusejí být formálně stanovena“ (ČSN ISO 5127-1/1.1.1-02)
- Umělý - „jazyk vytvořený nebo řízený pomocí souboru předem stanovených pravidel“ (ČSN ISO 5127-1/1.1.1-03)

Vlastnosti přirozeného jazyka

- Jazykové znaky lze rozložit na zvukové prvky, které jako znaky samy nemohou fungovat (subznaky), ale jejichž kombinací vzniká znak.
- Jazykové znaky je možné spojovat ve větší celky. Subznaků, které nenesou význam, je velice málo (13 - 80). Elementárních znaků, které jsou složeny ze subznaků, je poměrně mnoho, ale jejich počet je přece jen omezený. Z nich je však možné skládat nekonečné množství různých výpovědí (superznaků).
- Arbitrérnost (libovolnost) znaků je v případě jazyka omezená. V řeči se objevují slova zvukomalebná (ikonická) nebo je forma znaku motivována přímo vztahem k jiným znakům.
- Jelikož je jazyk základním nástrojem k uchopování světa, s narůstajícím poznáním některé znaky přibývají a jiné mizí – jazyk je tedy otevřeným znakovým systémem.

Srovnání PJ x SJ

Srovnání:

přirozený jazyk

X

selekční jazyk

asymetrický

- jeden pojem více slovy
- jedno slovo více pojmů

symetrický

- každému významu přiřadit jeden symbol

apelační + komunikační f-ce

- nařít, emoce, přesvědčit

jen komunikační f-ce

asociační spojení skrytá

asoc. spojení lze vyjádřit zjevně - zvýší se síla jazyka

možnost obměny slov nebo vyjádření

je silně omezena a nežádoucí, např. management X podniky -- řízení; soutěž hospodářská -- strategie X strategie konkurenční

Srovnání PJ x SJ - pokračování

ústní užití - je brán ohled na fonetiku
(d X t) záznamy

slova libovolně dlouhá

je redundantní (nadbytečný)

dominuje podst.jméno a sloveso

slovesné kategorie

dovede popsat sama sebe (metajazyk)

fonetika nezávažná, protože se hledají písemné

mnohdy je rovnoměrnost kódu pro slova žádoucí

omezujeme redundanci, např. jídla -- příprava

dominuje podst. jméno a přídavné jméno

nemají slovesné kategorie

SJ se popisují přirozeným jazykem

pravidla se tvoří před použitím

Terminologie

- Lexikum – slovní jednotka
- Lexikální jednotka – minimální posloupnost znaků, která nese význam

Selekční jazyky

- **Pořadací znak**
- **Gramatika** – pravidla tvorby tvarů pořadacích znaků a jejich spojování do vyšších celků
- **Morfologie** – část gramatiky zabývající se vytvářením tvarů pořadacích znaků
- **Syntax** – část gramatiky zabývající se spojováním pořadacích znaků

Systematické selekční jazyky

- **Systematické selekční jazyky**
- **Klasifikační znak** vyjádřen **notací**
- **Třída** – skupina klasifikovaných pojmů se shodnými vlastnostmi (vyjádřena notací)

Notace

- Podle znaků
 - Alfabetická
 - Numerická (DDC)
 - Alfnumerická (LCC)
- Podle struktury
 - Expanzivní
 - Lineární – vyjadřuje pořadí tříd nikoliv však vztahy mezi nimi
 - Hierarchická – vyjadřuje vztahy mezi třídami

Předmětové selekční jazyky

- Lexikální jednotka
- U tezurů
 - Deskriptor (preferovaný termín)
 - Nedeskriptor (nepreferovaný termín)
- Předmětové heslo

Jazykové jevy

- **Homonymie** - „vlastnost dvou nebo více termínů, které mají stejnou grafickou nebo zvukovou podobu, ale rozdílný význam“ (ČSN ISO 5127-1/1.1.2-16)
- **Homografie** - „vlastnost dvou nebo více termínů, které mají stejnou grafickou formu, ale rozdílný význam“ (ČSN ISO 5127-1/1.1.2-16)
- **Polysémie** - „vlastnost slova, které má dva nebo více etymologicky příbuzných významů“
- **Synonymie** - „vlastnost dvou nebo více termínů majících odlišnou formu a přesně nebo přibližně tentýž význam“

Vztahy v selekčních jazycích

- **Hierarchický vztah** - „formální vztah mezi dvěma termíny nebo třídami, kde jeden (jedna) je podřízen (podřízena) druhému (druhé)“ (ČSN ISO 5127-6/3.4.4-09).
- **Asociativní vztah** - „sémantický vztah mezi pojmy se vzájemnou vazbou z hlediska specifického účelu“ (ČSN ISO 5127-6/3.4.4-04).
- **Ekvivalence** - „formální vztah mezi termíny představovanými stejným (stejnými) deskriptorem (deskriptory) nebo znakem třídy“ (ČSN ISO 5127-6/3.4.4-11).

Procesy pořádání informací

- Proces vstupního zpracování informací v informačním systému
- Předchází informační analýza, zjišťují se významné identifikační a obsahové charakteristiky
- **Identifikační pořádání** – formální charakteristiky
- **Věcné pořádání** – obsahové charakteristiky
 - Systematické – systém poznání, rodo-druhov \acute{e} vztahy, notace
 - Předmětové – soubor abecedně uspořádaných hesel, přirozený jazyk, nevhodné pro klasifikaci

Pořádání

- **Indexace** - „proces vyjádření výsledku analýzy dokumentu prostřednictvím prvků selekčního jazyka nebo přirozeného jazyka, obvykle s cílem umožnit zpětné vyhledávání“
 - **Automatická**
 - **Polo-automatická**
 - **Intelektuální**
- **Post-kordinovaná** – bez předem stanoveného uspořádání znaků, ke koordinaci dochází až při vyhledávání
- **Pre-koordinovaná** – uspořádání dáno selekčním jazykem
- **Klasifikace** – přidělování notací (tříd) za účelem vystižení obsahu dokumentu
- **Fazetace** – rozdělení slovníku selekčního jazyka do obecných kategorií, vlastnosti pro seskupování pojmů podle jejich podstaty

Informační jazyk

- Jazyk užívaný v informačních systémech, dělí se:
 - Algoritmické informační jazyky
 - Logické informační jazyky
 - Selekční informační jazyky – SL je „formalizovaný jazyk používaný k charakterizování dat nebo obsahu dokumentů za účelem jejich ukládání a vyhledávání“
 - Identifikační selekční informační jazyky
 - Věcné selekční identifikační jazyky
- Selekční obraz dokumentu – množina všech pořadacích znaků přidělených dokumentu

Věcné selekční jazyky

- **Systematický selekční jazyk** – selekční jazyk používaný pro „pro strukturní zpracování dokumentů nebo dat pomocí symbolů a příslušných termínů s cílem umožnit systematický přístup, v případě potřeby s pomocí abecedního rejstříku“ (ČSN ISO 5127-6/3.4.1-03)
- **Předmětový selekční jazyk** - selekční jazyk používaný pro strukturní zpracování dokumentů nebo dat pomocí abecedně uspořádaných termínů s cílem umožnit předmětový přístup
- **Prekoordinovaný selekční jazyk** je selekční jazyk, jehož lexikum se skládá¹⁰ z pořadacích znaků, které vyjadřují složené pojmy a které jsou používány pro indexaci i vyhledávání.
- **Postkoordinovaný selekční jazyk** je selekční jazyk, jehož lexikum se skládá¹¹ z pořadacích znaků, jež vyjadřují jednoduché pojmy, při indexaci jsou do selekčního obrazu dokumentu zařazovány nezávisle na sobě a k jejich kombinaci dochází až v průběhu vyhledávání.

Selekční jazyky podle šíře

- Univerzální
 - Polytematické
- Speciální
 - Oborové

Systematické selekční jazyky

- Pre-koordinované (MDT, DDC, LCC)
- Post-koordinované (Ranganathan)

- Enumerativní – MDT, DDC, LCC
- Fasetové (analyticko-syntetické, Ranganathan, 44 faset – žánr + PMEST)

Předmětové selekční jazyky

- **Klíčová slova** – užívají slov z dokumentů, rejstříky KWIC a KWOC
- **Předmětová hesla** – jednoduché pořadací znaky, složky – heslo, podheslo, doplněk, obvykle pre-koordinované
- **Deskriptorové selekční jazyky** - význam a struktura jednotek specificky vymezena, post-koordinovaný jazyk, deskriptory, unitermy

Předmětová hesla

- osobu: Maciuchová, Hana, 1945-
- korporaci: Moravské divadlo Olomouc
Flora Olomouc 2004
Kostel Nejsvětější Trojice (Písek, Česko)
- rod, rodina: Slavníkovci (rod)
Krylovi (rodina)
- název dokumentu: Bible
Rukopis zelenohorský
Amsterodamská smlouva, 1997
Versailleská mírová smlouva, 1919
- autor / název: Kafka, Franz, 1883-1924. Proces
Aristotelés, 384-322 př. Kr. Politica
- věcné téma: psi
pěstování okrasných rostlin
prusko-rakouská válka, 1866
Lidice, vyhlazení, 1942
Srpen, 1968
dům čp.17 (Dobruška, Česko)
Československo -- národní hospodářství -- 1989-1992
etika -- díla do r. 500
Švejk, Josef (literární postava)
- geografický název: Hodolany (Olomouc, Česko)
Jávské moře
Morava, řeka
- forma/žánr dokumentu: učebnice základních škol
české romány
vědecko-fantastické povídky
terminologické slovníky

Deskriptorové selekční jazyky

- **Uniterm** - „nejmenší významový prvek selekčního jazyka použitý k vyjádření specifického pojmu v rámci systému koordinovaného indexování“ – většinou jednoslovné, odstraňuje homonymii a synonymii, není hierarchie
- **Deskriptory** – Tezaurus - „slovník řízeného selekčního jazyka uspořádaný tak, že explicitně zachycuje apriorní vztahy mezi pojmy“ (SY, BT, NT, RT)
 - Tezaury – univerzální, speciální, polytematické, oborové, jednojazyčné, vícejazyčné