**MATEMATIKA A STATISTIKA**

**1. Logika a její aplikace na přirozený jazyk.**

* **logika** – tj. věda o správném usuzování (způsobu vyvozování závěrů); jestliže usuzujeme a argumentujeme, používáme jazyk. Abychom mohli správné usuzovat a argumentovat, musíme rozumět jednotl. výrazům a větám (tzn. znát jejich význam); aby bylo možné reprezentovat význam vět s přihlédnutím ke znalostem o světě, kt. v systém pro ZPJ zachyceny pomocí speciální reprezentace znalostí –> např. systém založené na PK1 // TIL
* ***logická analýza PJ*** – analýza významů výrazů (vět) PJ; princip: přepsat PJ do formální logiky (tj. snaží se popsat strukturu jazykových výrazu pomocí logiky, tj. převést je formální logiky, a využít je pro dedukci (logické vyplývání), logická analýza má ***3 stupně***: 1. pomocí výrokové logiky (kdy vlastně ani nedochází k analýze jako takové, ale jde spíš analýzu logických spojek), 2. stupeň je výroková logika (logika neanalyzovaných pojmů), 3. logika analyzovaných pojmů (TIL, Montagueho intenzionální logika); jedná se v zásadě o nějakou kombinaci sémantické a syntaktického přístupu; výsledkem logické analýzy jména je termín, věty logická formule; zahrnuje analýzu pravdivostních vztahů mezi výroky (tj. analýza logického vyplývání)
* ***textové vyplývání***: 2 úseky textu –> úkolem je rozhodnout, zda význam 1 textu vyplývá (je odvoditelný) z 2. (soutěž od r. 2004)
* využití logické analýzy PJ: tvorba IS (porozumět a vyhovět uživateli, tzn. analyzovat jeho výroky) – složitější dialogové systémy
* intenze (tj. objekty závislé na čase a světě) x extenze (objekty na čase a světě nezávislé)
* ***TIL*** (Transparentní Intenzionální Logika) – autorem Pavel Tichý, tj. logický systém speciálně navržený pro zachycení významů PJ, obdoba Intenzionální logiky Montagueho – Tichý ukazuje její nedostatky; využívá výsledků teorie typů a intenzionální logiky
* vlastnosti: (1)
* ***nevhodnost výrokové logiky***: příliš nízká expresivita (na rozdíl od PJ)

a) TIL je logick ̋ systém zaloæen ̋ na urËité modifikaci (viz zejména dále pod b)) ty- povaného lambda kalkulu. Lambda kalkul je logick ̋ aparát, kter ̋ umoæÚuje manipulaci s funkcemi. Rozumná interpretace tohoto aparátu, kter ̋ má obecnÏ velké uplatnÏní v matematice a informatice, je umoænÏna principem teorie typ ̆, kter ̋ tvorbu funkcí omezuje na základÏ v ̋stavby tzv. hierarchie typ ̆ a podle nÏhoæ funkce nem ̆æe b ̋t apli- kována nap ̄. na sebe samu. Typovan ̋ lambda kalkul manipuluje s funkcemi v souladu s principem teorie typ ̆. Tím, æe je zaloæen na neomezené hierarchii typ ̆, je typovan ̋ lambda kalkul vhodn ̋m aparátem k p ̄ekonání nedostateËné expresivity, jaká je vlastní nap ̄. PL1.

I jiné systémy neæ til, zejména jiné intenzionální logiky, jsou zaloæeny na aparátu typovaného lambda kalkulu. Pokud vπak modifikují tento aparát, pak nikdy ve smyslu b), resp. c) (viz dále).

2V následujícím v ̋kladu se budeme opírat o ̄adu formulací z této práce. PodrobnÏjπí charakteristika formálního aparátu til je uvedena v p ̄íloze v odd. 0.9.5



75

b) TIL je transparentní systém, tj. pro til není formální aparát reprezentující zp ̆soby, jak ̋mi jsou konstruovány objekty, p ̄edmÏtem studia, n ̋bræ pouze prost ̄edkem ke studiu tÏchto konstrukcí .

Tímto rysem se til odliπuje od vπech soudob ̋ch logick ̋ch systém ̆: zatímco v til je formální v ̋raz oznaËením konstrukce, je pro stoupence formalismu tento v ̋raz bez- prost ̄edním jménem konstruovaného objektu. Na triviálním p ̄íkladu lze tento rozdíl ukázat takto:

formální pojetí TIL

------------------------------------------------------------------------

v ̋raz 3+5

sémantika Ëíslo 3 Ëíslo 5

sloæek

v ̋razu

operace sËítání

sémantika Ëíslo 8

v ̋razu

3+5 Ëíslo 3 Ëíslo 5

operace sËítání

konstrukce, tj.urËit ̋ zp ̆sob,

jak ̋m uvedené sloæky spolu-

pracují na vytvo ̄ení objektu

------------------------------------------------------------------------

Vidíme, æe pro formalistu neexistuje sémantick ̋ mezistupeÚ mezi objekty oznaËen ̋mi sloækami sloæeného v ̋razu a objektem v ̋sledn ̋m. Pro til je sémantika v ̋razu dána tím, æe zp ̆sob, jak ̋m je tento v ̋raz strukturován, zobrazuje strukturu konstrukce, jejímiæ sloækami nejsou sloæky jazykového v ̋razu, n ̋bræ objekty tÏmito sloækami oznaËené. Jak ukázal autor til v ̄adÏ statí (a zejména ve své monografii, Tich ̋, 1990), vede ignorování pojmu konstrukce k ̄adÏ chyb, nedorozumÏní i pseudoproblém ̆.

c) TIL nepreferuje jistá vybraná slova jako tzv. logická slova , jeæ by údajnÏ urËovala charakter logiky.

Také tento rys je specifick ̋ pouze pro til (souvisí s rysem b)). V ostatních, for- málnÏ budovan ̋ch systémech se vædy setkáváme s mnoæinou vyËlenÏn ̋ch konstantních v ̋raz ̆, které jsou logické a které jedinÏ zajiπaují odliπení logicky pravdiv ̋ch vÏt, logic- kého vypl ̋vání, logické ekvivalence od ostatních (z ̄ejmÏ na empirii závisl ̋ch) vlastností a vztah ̆. Tak ve v ̋rokové logice jsou logick ̋mi slovy logické (v ̋rokové) spojky, v PL1 k nim p ̄istupují kvantifikátory, resp. identita. Tato logická slova jsou navíc chápána jako tzv.nevlastní symboly, tj.interpretací jim není p ̄i ̄azován sobÏstaËn ̋ v ̋znam; v ̋znam je p ̄i ̄azován jen cel ̋m sloæen ̋m v ̋raz ̆m, které je obsahují.

Z tohoto hlediska nap ̄. vÏta (15) Pavel je starπí neæ Petr.

76

není logicky ekvivalentní vÏtÏ (16) Petr je mladπí neæ Pavel.,

protoæe anal ̋za tÏchto vÏt v PL1 dává (15’) St(Pavel, Petr), resp.

(16’) Ml(Petr, Pavel),

takæe se nem ̆æeme op ̄ít o æádné logické slovo, na jehoæ základÏ bychom mohli odvodit ekvivalenci (15) a (16). Samoz ̄ejmÏ, i PL1 odhalí logickou souvislost tÏchto vÏt tím, æe zavede v ̋znamov ̋ postulát

(17) 8 xy (St(x,y) ¥ Ml(y,x))

a prohlásí, æe (15’) je ekvivalentní s (16’) za p ̄edpokladu (17). Ale (17) je z hlediska intuice logicky pravdivá vÏta, takæe ji nepokládáme za zvláπtní p ̄edpoklad. Jenæe (17) nem ̆æe b ̋t z hlediska PL1 logicky pravdivá vÏta: aby jí byla, musela by b ̋t pravdivá ve vπech strukturách. Snadno vπak najdeme takovou strukturu, v níæ (17) neplatí; staËí za U zvolit nap ̄. mnoæinu p ̄irozen ̋ch Ëísel a za relace, jeæ budou interpretací p ̄i ̄azeny St, resp. Ml, relace >, resp. ∏.

Dalπí charakteristiky til se t ̋kají aplikace til na anal ̋zu p ̄irozeného jazyka. d) TIL aplikována na anal ̋zu p ̄irozeného jazyka se stává sémantikou zaloæenou na pojmu

moæn ̋ch svÏt ̆ (possible worlds semantics).

Tento rys sdílí til s nejrozπí ̄enÏjπími aplikacemi logick ̋ch systém ̆ na anal ̋zu p ̄irozeného jazyka. Myπlenka vyuæít moæn ̋ch stav ̆ svÏta, pop ̄. Ëasov ̋ch okamæik ̆ k de- finování intenzí jako logicky manipulovateln ̋ch objekt ̆ se stala v soudobé logické sé- mantice p ̄evládající ideou.

Poznámka:

Termín moæn ̋ svÏt byl p ̄evzat z Leibnize a poprvé v zárodeËné moderní podobÏ pouæit R.Carnapem. NÏkdy se mluví i o mnoæinÏ index ̆ (Montague aj.), do níæ jsou vedle moæ- n ̋ch svÏt ̆ a Ëasov ̋ch okamæik ̆ za ̄azovány nÏkteré dalπí parametry (ponejvíce pragmatické povahy). S kategorií moæn ̋ch svÏt ̆ pracuje i tzv.finská logická πkola (J. Hintikka aj.).

e) Univerzum je v TIL chápáno jako mnoæina spoleËná vπem moæn ̋m svÏt ̆m. Tento rys je charakteristick ̋ zejména pro til; ve vÏtπinÏ ostatních koncepcí se uvaæuje vedle moæn ̋ch svÏt ̆ i o moæn ̋ch individuích, tj.populace individuí je obecnÏ r ̆zná v r ̆zn ̋ch moæn ̋ch svÏtech. Tento zdánlivÏ samoz ̄ejm ̋ p ̄edpoklad (v nÏkterém moæném svÏtÏ existuje Pegas, v jiném ne) byl koncepcí til p ̄esvÏdËivÏ vyvrácen.

f) Fregeho (Churchovo) rozliπení vztahu denotace jakoæto oznaËování (reference) a vztahu vyjad ̄ování smyslu je v TIL zruπeno a nahrazeno jin ̋m schématem.

Také tento rys nalezneme u malého poËtu jin ̋ch systém ̆; vÏtπinou je denotace (oznaËení, pojmenování, reference) vztaæena k extenzím a intenze jsou chápány jako v ̋sledek zp ̆sobu vyjád ̄ení.

77

Vedle tÏchto rys ̆ charakteristick ̋ch pro til je t ̄eba se zmínit o specifickém de- duktivním aparátu, kter ̋ je obdobou syntaktického d ̆kazového aparátu v PL1, ale je p ̄izp ̆soben transparentní koncepci; neklade d ̆raz na axiómy, je generalizací Gentze- novy p ̄irozené dedukce (s touto teorií se lze seznámit nap ̄. v JanákovÏ práci, (1973)) na teorii typ ̆ a je velmi úËinn ̋. Nejjednoduππí aplikace tohoto aparátu byla u nás realizo- vána v systému ADAM pro reprezentaci znalostí na poËítaËi CYBER 172. (Viz T. Chrz, 1984).

0.7.2 Formální aparát – TIL a teorie typ ̆

P ̄edchozí úvahy nás vedou k hledání formálního aparátu vhodného pro sémantickou anal ̋zu v ̋raz ̆ PJ. Jak jsme uæ naznaËili, za takov ̋ nástroj pokládáme zmínÏn ̋ jiæ til.

Základními rysy systému til jsou:

* schopnost systematicky p ̄ekraËovat omezení platná v predikátové logice 1. ̄ádu  (extenzionální sémantice);
* d ̆sledn ̋ intenzionalismus a z nÏho vypl ̋vající schopnost p ̄esného definování in- tenzí a zacházení s nimi;
* vzhledem k p ̄irozenému jazyku disponuje til vÏtπí expresívní silou – coæ plyne z bodu 1.

PodrobnÏjπí charakteristiku systému til a jeho vlastností, díky nimæ je tak zajímav ̋ a vhodn ̋ pro sémantickou anal ̋zu PJ, uvádíme samostatnÏ v p ̄íloze Teorie typ ̆. I zde primárnÏ vycházíme z citované jiæ práce Materna, Pala, Zlatuπka, 1989.

**PALA str. 76**

**https://www.youtube.com/watch?v=j2eQ\_T7S-vw**

**2. Matematické objekty – množiny, relace, uspořádání, posloupnosti, ...**

**Množiny**

* ***matematický objekt*** – prvek množiny mající svůj název (název objektu) a symbol (symbol objektu, kt. jej zastupuje v symbolických zápisech) -> symbol objektu dvojího druhu: (a) konstanta (zastupuje konkrétní objekt z dané množiny objektů; 4, pí) a (b) proměnná (zastupuje libovolný objekt z množiny objektů, zpravidla písmena x, y)
* ***množina*** – tj. soubor objektů (nezáleží na tom, co tyto objekty představují), podstatné je, že objekty jsou od sebe různé a lze je vzájemně od sebe odlišit a lze jednoznačně určit, které objekty do množiny patří a které ne; utváří jeden celek; každý objekt patřící do dané množiny se nazývá **prvkem množiny**; není uspořádaná
* ***značení množin*:** pro množiny velká písmena latinské abc (A, B, M), pro prvky malá písmena (a, x, m)
* ***jazyk teorie množin***: jazyk predikátové logiky rozšířený o symboly {}, ∅, ∈ (množina je v tomto jazyce definována jako axiom); a ∈ A (a je prvkem množiny A), b není prvkem množiny A; mohutnost množiny – tj. počet prvků |A| = 3
* ***typy množin***: **neprázdná** (min. 1 prvek), **prázdná** (P= ∅; P = {}); **konečná** (počet prvků je určen přirozeným číslem); **nekonečná** množina; **potenční** (tj. množina všech podmnožin dané množiny; zápis: P(A) nebo 2A, tj. potenční množina P(A) množiny A je rovna množině všech podmnožin množiny, formálně P (A) = {B | B ⊆ A}, je vždy neprázdná, obsahuje totiž prázdnou množinu)
* ***zadání množin***: ***(a) výčtem*** (vyjmenováním) **prvků** – tj. uvedeme všechny prvky množiny (možné jen u konečných množin!); M = {m, n, p,...}, ***(b) logickou formulí*** (tj. výčtem charakteristických vlastností); M = {x ∈ Uł v(x)}: "M je množina všech x z množiny U, pro které platí v(x)"
* ***množinové vztahy***: **(1)** **inkluze** množin A a B (být podmnožinou je relace; A⊆B⇔∀x∈A:x∈B(pokud A je podmnožinou B, pak musí platit, že všechny prvky, kt. obsahuje množina A, musí obsahovat také množina B); **(2) rovnost** (dvě množiny jsou si rovny, je-li A podmnožinou B a B podmnožinou A; tzn. že mají stejné prvky, rovnost se dokazuje pomocí inkluze), popř. (3) ostrá inkluze (⊆ale pod tím je nerovná se): A je podmnožinou B, ale současně si nejsou rovny
* ***množinové operace***: **(1) sjednocení** (∪): sjednocením množin A a B vzniká nová množina obsahující všechny prvky z A i B (vyjma duplicit; vlastnosti: pokud sjednotíme dvě stejné množiny, dostaneme zase tutéž množinu; dále sjednocení je komutativní, nezáleží na pořadí a prázdná množina neobsahuje žádný prvek, takže není co sjednocovat), **(2) průnik** (***∩***): vzniká nová množina obsahující prvky, kt. mají tyto 2 původní množiny společné (náleží tam i tam) A***∩***B=∅, tzn. nemají žádný společný prvek, nazývají se A a B disjunktními množinami, **(3) rozdíl** (A \ B): nová množina obsahující prvky náležející množně A a současně nenáležející množině B; **doplněk množiny** (A') se zavádí v souvislosti s rozdílem množin, tj. všechny prvky, kt. nejsou v množině A.
* ***grafické znázornění množin***: množinové diagramy (konkrétně Vennovy diagramy – pomocí kt. se znázorňují množinové vztahy a operace) – užívají se, pokud chceme dokázat množinovou rovnost
* ***n-tice objektů*** (prvků): množina o n-prvcích (tedy konečná množina) – na pořadí prvků nezáleží (tj. neuspořádaná n-tice), anebo na jejich pořadí záleží (stanovíme, kt. prvek je 1./2. atd, tj. uspořádaná n-tice neboli konečná posloupnost; ***neuspořádaná dvojice***: {a, b}={b, a}je to tedy v zásadě množina, pokud a=b, pak je {a, b} jednoprvková, jinak dvouprvková; ***uspořádaná dvojice***: (a, b) – a je 1. prvek, b je 2. prvek, vyjádření pomocí neuspořádané dvojice {{a}, {a, b}}
* ***kartézský součin množin*** (A x B): kartézský součin 2 množin (A, B) je množina všech uspořádaných dvojic, v kt. je 1. prvek množiny A a 2. prvek z množiny B
* ***množinová zobrazení*** (fce; automat na kávu) – ve fci je každé vstupní hodnotě x jednoznačně přiřazena výstupní hodnota y, jedná se o relace; binární relace f na množině A je fce (unární fce je binární relace; binární fce je ternární relace); A= definiční obor/vstup), B= obor hodnot/výstup; množina A je fcí právě tehdy, existuje-li ke každému prvku x z jejího definičního oboru jeden prvek z oboru hodnot
* ***vlastnosti fcí:*** ***(1) injenktivita*** (prostá fce): žádné 2 vzory (definiční obor) nemají stejný obraz (tj. existuje jiný obraz) (obor hodnot)(1 –> 1, 2 –> 3, nikoli pak 3 –> 3); ***(2) surjektivita*** (sirjektivita): všechny hodnoty z oboru hodnot mají nějaký vzor (clý obor hodnot je pokryt); ***(3) bijektivní***: fce je bijekcí, je-li úplná, injenktivní a surjektivní; ***úplná*** – pro každý vzor existuje obraz

**Relace**

* ***definice***: hlavní definice (tj. **množina uspořádaných dvojic/n-tic**): relace využíváme jako způsob, jak svázat nějaké 2 hodnoty (objekty) k sobě, tzn. chceme vyjádřit, že mají něco společného; relace modeluje vztah mezi objekty pomocí prostředků z teorie množin (z dvojic prvků, mezi nimiž existuje jistý vztah, se utvoří uspořádaná dvojice a z těchto uspořádaných dvojic se dále utvoří množina (neboli relace); často se pak říká "relace na množině A"
* ***binární relace***, n-ární relace
* ***vlastnosti binárních relací***: (***1) reflexivita*** (prvek je v relaci sám se sebou), ***(2) symetrie*** (a je v relaci b a a s b), ***(3) antisymetrie*** (pokud je a v relaci s b a b s a, pak a=b), ***(4) tranzitivita*** (a je s b, b je s c, pak a i c v relaci)
* ***2 zákl. typy relací***: ***relace ekvivalence*** (relace pro vyjádření stejnosti, podobnosti, izomorfismu; každá relace, kt. je symetrická, tranzitivní a reflexivní) a ***relace uspořádání*** (reflexivní, tranzitivní, antisymetrická)

**Posloupnosti**

* ***definice***: tj. fce, jejímž definičním oborem je množina přirozených čísel; konečná číselná posloupnost (uspořádaná n-tice; definičním oborem je množina prvních n přirozených čísel), nekonečná (definičním oborem je celá množina přirozených čísel)
* ***Fibonacciho posloupnost***: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34; a0= 0; a1= 1; an = an-1 + an-2

**3. Statistika a pravděpodobnost – statistický soubor a jeho charakteristiky, pravděpodobnostní prostor.**

**Statistika**

* ***statistika*** – (a) jako teorie: se zabývá metodami sběru, zpracování a vyhodnocení statistických údajů; (b) jako činnost znamená získávání statistických údajů (pozorováním, měřením, vážením apod.), jejich zpracováním a hodnocením, ***matematická statistika*** – vychází z již nahromaděných statistických údajů a zabývá se jejich matematickým (statistickým) zpracováním a rozborem získaných dat; statistická data – tj. kvantitativní (číselné) údaje o zkoumaných objektech
* ***statistický soubor*** – tj. množina/posloupnost údajů (zpravidla číselné povahy) o nějakých objektech (údaje zhromažděné na základě určitých společných vlastností těchto objektů) (***základní statistický soubor*** (populace) – zahrnuje všechny objekty stejného typu (všechny děti ze školek) –> není možné změřit a zvážit všechny děti, proto pracujeme pouze s určitým výběrem ze zákl. stat. souboru – tj. statistický soubor –> aby tento vzorek správně vypovídal o zákl. statist. souboru spoléháme se většinou na náhodný vzorek**)**
* ***vlastnosti statistického souboru***: typy údajů – **statistické znaky** (hodnoty znaku); jejich počet – **rozměr statistického souboru**; příklad: statistický soubor – množina údajů o výšce a hmotnosti dětí ve školce; výška+hmotnost – tj. statistické znaky; rozměr statist. souboru – 2, **statistické jednotky** (prvky) – tj. zkoumané objekty; **rozsah souboru** – tj. počet všech prvků

**příklad:** naměřili jsme 6 dětí s váhou 18, 20, 20, 20, 22 a 24 kg

* ***druhy statistických souborů***:

**(1) jednorozměrný**: statistický soubor omezený na jediný statistický znak – váha; rozsah statistického souboru: 6; ***absolutní četnost*** hodnoty znaku – počet jejich výskytů v souboru, tj četnost hodnoty 20 je 3; ***relativní četnost*** – tj. absolutní četnost podělená rozsahem souboru, tj. 50 %; ***kumulativní četnost***– tj. četnost hodnoty znaku + četnost všech nižších hodnot; pro hodnotu 4 je to 66 %; četnosti se zaznamenávají pomocí histogramu (sloupcový graf fce, kt. má na ose x jména tříd, na ose y jejich absolutní a relativní četnosti)

**Pravděpodobnost**

* pravděpodobnost jevu – tj. míra očekávání toho, že náhodný jev nastane
* ***náhodná proměnná*** – např. A je vlastnost, jejíž hodnotu neznáme (např. hod kostkou); ***pravděpodobnostní rozložení*** (je fce) udává, s jakou pravděpodobností nabude určitých hodnot, s tím, že hodnota pravděpodobnostního rozložení (p) je rovna pravděpodobnosti (P; obecná pravděpodobnost), dvojice (X, p) – tj. množina všech možných hodnot náhodné proměnné spolu s pravděpodobnostním rozložením – tj. ***pravděpodobnostní prostor***;
* ***určení pravděpodobnostního rozložení*** – zákl. metoda spočívá v určování pravděpodobnosti na základě dat stejného typu, jejichž měření bylo provedeno již v minulosti; ***3 typy*** pravděpodobnostního rozložení: **(1) uniformní** (všechny hodnoty mají přibližně stejnou pravděpodobnosti), **(2) normální rozložení** (nejpravděpodobnější hodnoty jsou blízko průměru, čím větší odchylka od průměru, tím pravděpodobnost klesá; graf tvaru zvonu), **(3) Zipfovo** (nejčastější hodnoty mají nejv. pravděpodobnost, s každou další hodnotou pravděpodobnost prudce klesá: x, x/2, x/3 ), platí **Zipfův zákon** – ten platí téměř všude v PJ (slova, sl. druhy, syntaktické vztahy apod.): že nejčastější slovo se vyskytuje 2x častěji než 2. nejčastější slovo –> vezmeme vždy frekvci slova x jeho pořadí = konstanta, konstanty všech slov jsou si podobné
* ***statistické testování hypotéz***: statisticky se prokazuje jistá hypotéza; ***nulová hypotéze*** (H0; snažíme se vyvrátit předpoklad, kt. většinou platí; ***alternativní hypotéza*** (H1; popírá platnost nulové hypotézy); ***chyba typu I*** (vyvrátíme H0, ale přitom platí, tj. potvrdíme H1); ***chyba typu II*** (nastává, když nepotvrdíme H1, ale ta je přitom platná)
* ***podmínění pravděpodobnost*** P(A|B) – tj. pravděpodobnost jevu A za předpokladu, že nastal jev B

**4. Aplikace statistiky na zpracování jazyka.**

* aplikace statistiky při automatickém zpracování PJ – víme-li, že statistika je nástroj, kt. umožňuje získat relevnantní info z velkého množství dat; chceme-lip statisticky zpracovat PJ, slouží nám k tomu korpusy, rozsáhlé soubory jazykových dat.
* ***kolokace*** – různé definice, pro potřeby statistiky je to statisticky významné spojení 2 slov (jak silná je kolokace – tj. spojení 2 slov);
* ***jak hledat v korpusu kolokace***? ***(1) T-test*** (aplikace testování hypotéz, kdy nulová hypotéza by byla, že slova se chovají náhodně dle svých pravděpodobnostních rozložení; pokud by pak byla nulová hypotéza vyvrácena – tj. slova se spolu vyskytují častěji než náhodně, pak prohlásíme příslušnou dvojici za kolokaci, za pravděpodobnosti chyby typu I pak pak můžeme odvodit sílu kolokace; ***(2) asociační míry*** – (zpravidla omezeny na bigramy, založeny na statistickém testování hypotéz), pro vyhledávání kolokací v korpusu, typicky pracují s frekvencí celé kolokace, jejich jednotl. slov, z kt. se skládá a s velikostí korpusu, dosazují se do kontigenční tabulky (vizualizace vzájemného vztahu 2 statistických znaků) a na základě vzorce vyjde určité číslo, toto číslo udává míru asociace (může být i záporné – slova se odpuzují); typy asociačních měr: **T-score** (testem náhodného rozložení; vychází ze statistického testování hypotéz pomocí tzv. t-testu; někdy označení jako míra kotrastu; nevýhoda: ovlivněn frekvencí celé kolokace) **MI-score** (výsledkem je hodnota/číslo, kt. udává pravděpodobnost toho, že 2 slova se vyskytnou současně vedle sebe; nevýhoda: velmi ovlivněn frekvencí jednotl. slov (nejvyšších hodnot pak dosahují spojení s nízkou frekvencí –> nastavuje se spodní hranice frekvence): ***Dice a logDice***, kt. se užívá pro sestavování kolokačních profilů (tzv. word sketches; vyhledá pro základní tvar slova, např. "voda"/"začít", jeho nejčastější vazby zachycující kolokační a gramatické chování tohoto slova), na rozdíl od ostatních měr počítá pouze s frekvencí slov a frekvencí celé kolokace (tj. bez celkové velikosti korpusu); nižšší T-score MI-score –> volné kombinace; vyšší –> těsné kombinace, vysoké (obvykle nad 10) –> kolokace
* ***frekvence*** (četnost) – v rámci celého korpusu/konkordance (hledané slovo + jeho kontext
* n-gramové jazykové modely

**5. Statistický jazykový (n-gramový) model – konstrukce, využití.**

* ***n-gram*** – tj. posloupnost n-jednotek stejného druhu (písmen, častěji však slov) v textu; bigram se zpravidla ztotožňuje s kolokací (ačkoli ty mohou více než dvouslovné a mezi členy bigramu nemusí být žádná provázanost či ustálenost); každá kolokace je bigramem, né však každý bigram je kolokací (např. bigram "jak se")
* trigramy, tetragramy apod. (zdálo se mi že, to je v pořádku)
* ***n-gramový jazyk. model*** – formálně se skládá z podmíněných pravděpodobností P (wn|w1,...wn-1) neboli pravděpodobnost, že se vyskytlo slovo wn, za předpokladu, že se před ním vyskytla slova w1 až wn-1; načež w1 až wn1 nemusejí být slova, může se vytvořit n-gramový model znaků, pádů, morfémů atd.; při zpracování PJ je nepoužívanější trigramový jazykový model; obecně umožňuje odhadnout nejpravděpodobnější sekvenci symbolů, kt. odpovídají (víceznačnému) vstupu
* ***využití***: zpracování PJ: rozpoznávání řeči, morfologické značkování, syntaktická analýza, strojový překlad

**6. Vyhodnocování aplikací zpracování jazyka – přesnost, pokrytí, F-míry a zlatý standard, křížová validace.**

* ***f-míra*** – je kombinací míry přenosti a míry pokrytí; počítá se tzv. s váhovým parametrem alfa, kt. se pohybuje od 0 do 1, čím vyšší je alfa, tím vyšší je míra přesnosti na úkor pokrytí apod, alfa=0,5 (tj. harmonický průměr přenosti a pokrytí; mají stejnou váhu)
* ***přesnost*** – míra vyjadřující, jak velká část výsledků je relevantní
* ***pokrytí*** – míra vyjadřující, jak velká část výsledků byla nalezena
* žádná měření nejsou 100 % – jde o to, aby se pohybovaly v určitém tolerančním poli
* ***křížová validace*** – pro ověření, jak funguje statistický model na různých datech; data se rozdělí na trénovací data a data na natrénování modelu

**1. Logika a její aplikace na přirozený jazyk**

Logická analýza PJ

#### Formální aparát pro SR – charakteristika TIL

**Transparentní intenzionální logika** – logický systém založený na modifikaci typovaného lambda kalkulu = logiký aparát umožňující manipulaci s funkcemi. Je založen na principu teorie typů. Je vhodný pro sémantickou analýzu výrazů PJ. Byl navržen speciálně pro zachycení významu výrazů PJ.

Pro TIL je sémantika výrazu dána tím, že způsob, jakým je tento výraz strukturován, zobrazuje strukturu konstrukce, jejímiž složkami nejsou složky jazykového výrazu, nýbrž objekty těmito složkami označené. (Např. ve formálním pojetí je sémantikou výrazu *3+5* číslo 8, v pojetí TIL je to určitý způsob, jakým uvedené složky spolupracují na vytvoření objektu.

Neobsahuje žádná „logická slova“ jako jsou výrokové spojky nebo kvantifikátory. TIL aplikována na analýzu přirozeného jazyka se stává sémantikou založenou na pojmu možných světů. Univerzum je chápáno jako množina společná všem možným světům. Vztah označování a vztah vyjadřování je nahrazen jiným schématem.

Zákldnímy rysy TIL systému jsou:

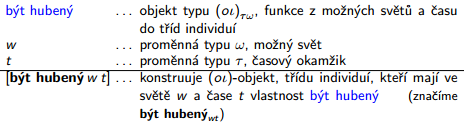
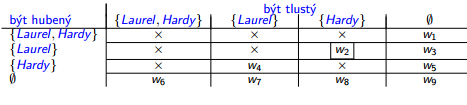
* schopnost systematicky překračovat omezení platná v predikátové logice 1. řádu
* intezionalizmus a z něho vyplývající schopnost přesného definování intenzí[[1]](#footnote-2) a zacházení s nimi
* větší expresivní síla
* má rozvětvenou typovou hierarchii
* temporální
* transparentní – nositel významu (konstrukce) není prvek formálního aparátu, tento aparát pouze studuje konstrukce, zachycení intenzionality je přesně popsáno z matematického hlediska

**Typy v TILu**

* základní typy – typová báze = {ο, ι, τ, ω}, umožňují přiřadit typ objektům z intenzionální báze jazyka – třída zákl. vlastností (barvy, rozměry, postoje, …) popisujících stav světa
  + ο (omikron) – pravdivostní hodnoty Pravda (true, T) a Nepravda (false, F), odpovídají běžným logikám
  + ι (jota) – třída individuí, individua jako numerická identifikace nestrukturované entity
  + τ (tau) – třída časových okamžiků (časová kontinua), zachycení závislosti na čase, třída reálných čísel
  + ω (omega) – třída možných světů, zachycení empirické závislosti na stavu světa
* funkcionální typy – funkce nad typovou bází
* typy vyšších řádů

**Možné světy**

* termín možný svět – Gottfried Wilhelm von Leibniz (1646 – 1716, ﬁlozof a matematik)
* požadavky na deﬁnici “možného světa:”
  + soubor myslitelných faktů
  + je konzistentní a maximální ze všech takových souborů
  + je objektivní (nezávislý na individuálním názoru)
* mezi možnými světy existuje právě jeden aktuální svět – jeho znalost ≡ vševědoucnost
* možný svět v TILu = rozhodovací systém, pro každý prvek intenzionální báze obsahuje konzistentní přiřazení hodnot, např. realita s 2 objekty a 2 vlastnostmi (9 možných světů):



**2. Matematické objekty – množiny, relace, uspořádání, posloupnosti,...**

**MNOŽINY, USPOŘÁDANÉ MNOŽINY A OPERACE NA NICH**

***Teorie množin***

* spolu s logikou zákl. pilíř matematiky
* všechny matematické objekty jsou množiny
* různé formální teorie (nekonečno, axiom výběru)
* např. Cantorova (Georg Cantor; 2. pol. 19 st.) – „Naivní teorie množin“: Množinou rozumíme souhrn určitých a rozlišitelných objektů existujících v naší mysli. Těmto objektům říkáme prvky množiny –> vyvrácena Russlem (holič)

***Množiny***

* tj. skupina objektů (čísel, aut..) – není důležité, co přesně tyto objekty (chceme-li prvky) jsou/co představují, důležité je, aby bylo možné jednotl. prvky od sebe odlišit a jednoznačně určit, kt. do množiny patří a které ne
* né nutně stejného typu
* neobsahuje duplicity
* není uspořádaná
* počet prvků v množině – tj. mohutnost množiny A zapisujeme: |A| = 2 (A = {1, 2})

***Zákl. fakta***

* existuje prázdná množina – P = {}, P = ∅
* množina může obsahovat množiny jiné

***Jazyk teorie množin***

* jazyk predikát. logiky rozšířený o symboly {,}, ∅, |, ∈
* pojem množiny je definován axiomy zapsanými v tomto jazyce

***Nekonečné množiny***

* zápis množin:
  + výčtem prvků: {1, 2, 3}, {∅, ∅}
  + logickou formulí: {x | x ∈N ∧ x > 5} –> tj. pomocí vlastnosti, kt. charakterizuje její prvky
    - množina všech sudých přirozených čísel: {m | m = 2k, k je přirozené číslo}.
* nekonečné množiny
  + existují ve většině teorií množin
  + různě velká nekonečna
  + např. přirozená čísla (racionální čísla) vs. reálná čísla
* jiné zadávání množiny – pomocí vlastnosti, kt. charakterizuje její prvky

***Podmnožina ⊆***

* být podmnožinou je relace
* formální definice: A⊆B⇔∀x∈A:x∈B(pokud A je podmnožinou B, pak musí platit, že všechny prvky, kt. obsahuje množina A, musí obsahovat také množina B
* podmnožinu můžeme též označit termínem inkluze

***Potenční množina***

* tj. množina všech podmnožin dané množiny
* zápis: P(A) nebo 2A –> potenční množina P(A) množiny A je rovna množině všech podmnožin množiny
* formálně: P (A) = {B | B ⊆ A}.
* je vždy neprázdná, obsahuje totiž prázdnou množinu

***MNOŽINOVÉ OPERACE***

***Operátor ∈***

* tzn. prvek patří do množiny (na levé straně je vždy prvek, na pravé množina)
* tj. stylizované řecké písmeno epsilon, „x ∈A” – tj. x (je) prvkem A

***Rovnost množin =***

* 2 množiny jsou si rovny, právě když A je podmnožinou B a B je podmnožinou A, neboli: A=B⇔(A⊆B ∧ B⊆A)

(tzn. že 2 množiny jsou si rovny, mají-li stejné prvky), důkaz rovnosti – 2 části, kt. se dokáží inkluzí

***Sjednocení množin ∪***

* sjednocením množin A a B vznikne nová množina obsahující všechny prvky jak z A, tak z B (vyjma duplicit)
* A∪B={x|x∈A ∨ x∈B}
* ***vlastnosti***:
  + *A*∪*A*=*A* : pokud sjednotíme dvě stejné množiny, dostaneme zase tutéž množinu
  + *A*∪*B*=*B*∪*A* : sjednocení je komutativní, nezáleží na pořadí.
  + *A*∪∅=*A* : prázdná množina neobsahuje žádný prvek, takže není co sjednocovat

***Průnik množin ∩***

* vzniká nová množina, kt. obsahuje prvky, kt. mají tyto 2 množiny společné (tj. množina obsahující prvky, kt. náležejí jak do množiny A, tak do množiny B)
* A∩B={x|x∈A ∧ x∈B}

**FCE A RELACE, JEJICH UZÁVĚRY**

***Neuspořádaná dvojice***

* je to vlastně množina {a, b} = {b, a} –> nezáleží na pořadí – třeba rozlišit množinu {a, b} a dvojici (a, b)
* pokud jsou prvky a, b od sebe různé, tzn. a ≠ b, potom je množina {a, b} dvouprvková; jsou-li a = b, pak je množina {a, b} jednoprvková
* tzn. že {x, y} = {y, x}, pokud x=y, pak je tato množina jednoprvková

***Uspořádaná dvojice – odvozená struktura***

* (a, b) – má 1. a 2. prvek (na rozdíl od množiny záleži na pořadí)
* pro libovolné 4 prvky platí (a, b) = (c, d), právě když a = c a současné b = d
* ***definice pomocí množin***:
  + uspořádanou dvojici lze vyjádřit neuspořádanou dvojicí: (a, b) ≡  {{a}, {a, b}} –> takto jednoznačně rozlišíme 1. prvek
  + lze vidět, že pojem uspořádaná dvojice je vyjádřen pouze pomocí prostředků teorie množin –> uspořádaná dvojice je určitá množina obsahující své prvky jako množiny –> množina {a, b} určuje, z jakých prvků se daná uspořádaná dvojice skládá a množina {a} určuje, kt. prvek je první
* ***uspořádaná n-tice*** (kde n je přirozené číslo)
  + uspořádaná trojice prvků a, b, c rozumíme uspořádanou dvojici (a,b,c) ≡ (a,(b,c)), obdobně uspořádanou čtveřicí prvků a, b, c, d je (a, b, c, d) = (a, (b, c, d)), přičemž (b, c, d) = (b, (c, d))
  + uspořádaná n-tice prvků a1, a2, a3, a4 ... an rozumíme uspořádanou dvojici –> obecně (a1,a2,a3,...,an)≡***(a1,(a2,(a3,(...,an)...)))*** (funguje jen pro konečné n)
  + pojem uspořádané trojice vychází z uspořádané dvojice, neboť (a, (b, c)) je uspořádaná dvojice, kde prvek a je její 1. složka a (b, c) (tj. sama uspořádaná dvojice!) je její 2. složka –> obdobně uspořádaná čtveřice (a, b, c, d) je uspořádaná dvojice (a, (b, c, d)), v jejíž 2. složce je uspořádaná trojice
  + tedy obecně jakoukoli uspořádanou trojici (celkově n-tici) chápeme vždy jako uspořádanou dvojici (a1, (a2, (a3, (...an)..)))

***Kartézský součin***

* *kartézský součin 2 množin* A, B (A x B) je množina všech uspořádaných dvojic, v které je 1. prvek prvek množiny A a 2. prvek uspořádané dvojice je prvek množiny B, formálně zapsáno: AxB≡{(a,b)|a∈A ∧ b∈B} –> množina uspořádaných dvojic prvků z A a B; tj. (x, y), kde x ∈A, y ∈B
* obecně kartézský součin není komutativní (zaměnitelný) –> záleží na pořadí prvků, pro A ≠ B platí: A x B ≠ B x A; jestliže |A| = n, |B|= m, pak |A x B| = n.m
* příklad: určete kartézský součin množin A = {a, b, c} a B = {1, 2}

řešení: A x B = {(a, b)|a ∈A, b ∈B} = {(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)}

B x A = {(a, b)|a ∈B, b ∈A} = {(1, a), (2, a), (1, b), (2, b), (1, c), (2, c)}

* + - oba kartézské součiny jsou stejně mohutné (díky komutativitě násobění\*), obsahují však dvojice s prvky v opačném pořadí
    - \* komutativní zákon platící pro sčítání a násobení, tzn. y \* x = x \* y; A x B ≠ B x A
    - kartézský součin není komutativní a není asociativní ((x \* y) \* z = x \* (y \* z), tzn. A x (B x C) ≠ (A x B) x C; je však distributivní: x \* (y + z) = x \* y + x \* z; tzn. A x B x C = A x (B x C)
* *kartézský součin více množin*
  + analogicky – obsahuje uspořádané n-tice

***RELACE***

***Obecně:***

* vedle množin jsou dalšími důležitými zákl. “datovými typy” matematiky relace
* ***definice***: je to množina udpořádaných dvojic (jsou-li X, Y množiny, nazývá se libovolná podmnožina kartézského součinu X x Y relací mezi X a Y), v případě X = Y –> mluvíme o relaci na X, což je tedy libovolná podmnožina R ⊆ X2
* náležejí-li dvojice (x, y) relaci R, tj. (x,y) ∈R –> říkáme, že x a y jsou v relaci R, tedy xRy –> vždy je třeba uvést ohledně jaké množiny se relace týká (např. relace na R to byla jiná množina dvojic)
* ***motivace***: způsob, jak v matematice svázat 2 hodnoty (případně více); vyjadřujeme, že objekty (množiny) v realci mají něco společného
* ***arita*** – kolik je prvků v relaci (unární, binární, ternární..)
* relace modeluje vztah mezi prvky pomocí prostředků teorie množin (z dvojic prvků, mezi nimiž platí určitý vztah, se vytvoří uspořádané dvojice, z těchto uspořádaných dvojic se pak utvoří množina (relace)
* často říkáme “relace na množině A”
  + tzn. podmnožina součinu A x A
  + resp. A x A x ...x A

***Binární relace***:

* binární relace – pokud je složena z uspořádaných dvojic (dává do vztahu dvojice prvků)
* –> podmnožina kartézsk. součinu
* přehledný zápis binárních relací:
  + tabulkou
  + grafem

***N-ární relace***

* množina uspořádaných n-tic

***Relace – příklady***

* rovnost (=) a neostrá rovnost (>=) – tj. relace na množině N všech přirozených čísel; u rovnosti to jsou dvojice (1, 1), (2, 2), (3,3), druhá z dvojic (1, 1), (2, 1), (2, 2), ...

***Relace identity na množině A (identická relace–identita)***

* Id(A) – binární relace (jedná se o binární relace)
* Id(A) = {(a, a) ∈A x A | a∈A}
* tj. nejmenší relace ekvivalence (jedná se o nejmenší relaci splňující reflexivitu – Relace R musí obsahovat minimálně dvojice [x, x] pro všechny prvky, což je právě relace identity (dále je i tranzitivní, symetrická)

***Relace větší nebo rovno na přirozených číslech***

* >= (N) – binární relace
  + >=... tj. neostrá rovnost
* >= (N) = {(a, b) ∈N x N | b ⊆  a} – podle množinové konstrukce přirozených čísel

***Relace plus na přirozených číslech***

* +(N) – ternární relace
* +(N) = {(a, b, c) ∈N x N x N | a + b = c}
* a + b = c je jen jiný zápis pro (a, b, c) ∈+(N)

–> všechny operace na číslech jsou relace

***Vlastnosti binárních relací***

***Reflexivita***

* relace je reflexivní, pokud je každý prvek množiny v relaci sám se sebou
* ∀a∈A((a,a)∈R) neboli– jestliže pro každé x ∈X platí xRx

***Symetrie***

* relace je symetrická, pokud každý libovolný prvek a je v relaci s jiným libovolným prvkem b a současně prvek b je v relaci s prvkem a
* R(A) je symetrická právě tehdy když ∀a,b∈A((a,b)∈R⇒(b,a)∈R)

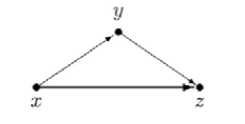


***Antisymetrie***

* relace je antisymetrická, pokud pro každý prvek relace platí, že pokud prvek a je v relaci s prvkem b a současně prvek b je v relaci s prvkem a, pak platí a = b a naopak
* R(A) je antisymetrická právě tehdy, když ∀a,b∈A((a,b)∈R∧(b,a)∈R⇒a=b)

***Tranzitivita***

* relace je tranzitivní, pokud pro každý libovolný prvek relace platí, že pokud prvek a je v relaci s prvkem b a současně prvek b je v relaci s prvkem c, pak prvek a je v relaci také s prvkem c
* R(A) je tranzitivní právě tehdy, když ∀a,b,c∈A((a,b)∈R∧(b,c)∈R⇒(a,c)∈R)
* podmínku trans(z)itivity lze dobře vyjádřit pomocí šipek: jsou-li šipky v relaci x –> y a y –> z, musí platit i x –> z



***2 zákl. typy relací: relace ekvivalence a relace uspořádání***

***Ekvivalence (relace ekvivalence)***

* pojem ekvivalence se užívá pro různé pojmy pro vyjádření stejnosti, podobnosti, isomorfismus apod.; relace ekvivalence se zpravidla značí =, ≡ apod.
* relace R na X je ekvivalence na X, jestliže je symetrická, tranzitivní a reflexivní (ekvivalencí nazýváme každou relaci, kt. je symetrická, tranzitivní a reflexivní)

***Uspořádání***

* uspořádáním nazýváme každou reflexivní, tranzitivní a antisymetrickou relaci

***Vlastnosti binárních relací – příklady***

***Identita na libovolné množině***

* splňuje všechny výše uvedené vlastnosti –> ekvivalence i uspořádání

***Relace ≤ na přirozených číslech***

* není symetrická –> uspořádání

***Relace < na přirozených číslech***

* není symetrická ani reflexivní –> ani ekvivalence, ani uspořádání

***Relace na prázdné množině R(***∅***)***

* je ∅ (podmnožina ∅ x ∅) → ekvivalence i uspořádání
* (pro všechny prvky prázdné množiny platí kde co)

***Ekvivalence a rozklad***

* každá ekvivalence na množině M indukuje rozklad, tj. množinu podmnožin (v nichž každý prvek je v relaci s každým, žádný prvek není v relaci s žádným prvkem z jiné podmnožiny –> tj. průnik libovolných 2 tříd je prázdný a jejích sjednocením je celá množina
* každá ekvivalence rozdělí množinu M na systém disjunktních množin, kt. se nazývají třídami ekvivalence) –> množiny jsou vzájemně disjunktní, tzn. že nemají společný žádný prvek, všechny prvky jsou však vzájemně ekvivalentní
* třída ekvivalence: Ax ={a∈A|(a,x)∈R} – tj. že všechna a jsou ekvivalentní se zadaným prvkem x (protože relace ekvivalence je reflexivní, tam se tímto způsobem dostane i samotný prvek x, protože x je ekvivalentní samo se sebou); rozkladem ekvivelance pak rozumíme množinu všech tříd ekvivalence
* rozklad množiny A podle ekvivalence R: A|R = {Ax |x∈A}

***Definice celých čísel***

* uvažujme množinu dvojic přirozených čísel D = {(a, b) ∈N x N}
* ... spolu s ekvivalencí R –> ((a, b,), (c, d)) ∈R ≡ a + d = b + c
* uvažujme rozklad podle této ekvivalence:
  + třídy rozkladu jsou Da,b = {(x, y) | ((x, y) | (x, y), (a, b)) ∈R}
  + rozklad D|R odpovídá množině {Da,b | a, b ∈N}
* tento rozklad je konstrukcí celých čísel Z
  + každá třída Da,b odpovídá číslu a–b

**FUNKCE (mezi množinami)**

* současné chápání – jako speciální typ relace; ve funkci f je každé vstupní hodnotě x přiřazena jednoznačně výstupní hodnota y
* tj. všechny fce jsou současné i relace (ale naopak to neplatí)
* fcím se také někdy říká zobrazení

***Alternativní pohled***

* prvních n-1 hodnot uspořádané dvojice jsou argumenty relace (vstup)
* posl. hodnota je hodnota fce (výstup)
* zápis – např.: (a, b, c) ∈+ ≡ + (a, b) = c

***Funkce***

* taková relace, kde výstup je jednoznačná
* binární relace f na množině A je fce, pokud platí:

∀a,b,c∈A((a,b)∈f ∧(a,c)∈f⇒b=c)  –> unární fce je binární relace

* ternární relace f na množine A je fce, pokud: ∀a,b,c,d ∈A((a,b,c)∈f ∧ (a,b,d)∈f ⇒c =d )

–> binární fce je ternární relace

* podobně fce proměnných

***Varianty zápisu***

* často říkáme “fce z A do B”
  + případně “zobrazení z A do B”
  + zapisujeme f: A –> B
  + tj. podmnožina kartézského součinu A x B
  + A = definiční obor (vstup), značíme Df nebo dom(f) – jedná se o unární operace (taktéž obor hodnot)
  + B = obor hodnot (výstup), značíme Rf nebo f(A)
  + pozn.: množina A je fce právě tehdy, když ke každému prvky x z jejího definičního oboru dom(A) existuje právě jeden prvek y z jejího oboru hodnot f(A) takový, že x, y ∈A
    - fce modeluje pojem zobrazení (a operace) pomocí prostředků teorie množin – ze vzorů a jejích obrazů vytvoříme uspořádané dvojice; z nich pak utvoříme množinu (fci)

***Fční hodnota***

* namísto standard. zápisu a, b ∈f, píšeme f(a) = b
* totéž jako: (a, b) ∈f
* také “b je obraz prvku a”
* také “a je vzor prvku b”

**VLASTNOSTI FCÍ**

***Injenktivita (neboli prostá fce)***

* f: A –> B je injenktivní (též prostá), právě tehdy, když ∀a, b ∈ A ( f (a) = f (b) ⇒ a = b ) –> žádné 2 nemají stejný obraz

***Surjektivita [syrjektivní]***

* f: A –> B je surjektivní (též „na“) právě tehdy, když ∀b∈B (∃a∈A(b=f(a))
* –> každý prvek oboru hodnot má nějaký vzor
* –> celý obor hodnot je pokrytý

***Úplnost***

* f: A –> B je úplná, právě tehdy, když ∀a∈A(∃b∈B (b=f(a))
* –> každý prvek definičního oboru má nějaký obraz
* –> celý definiční obor je pokrytý
* pojmem „fce“ se často myslí úplná fce

***Bijekce (bijektivní fce – neboli vzájemně jednoznačná fce)***

* f: A –> B je bijekce, právě tehdy, když je injektivní, surjektivní a úplná
  + - množiny A a B jsou stejně velké

***Inverzní fce***

* pokud f: A –> B je injektivní, definujeme inverzní fci
* f-1: B –> A
* f-1(b) = a ≡ f(a) = b

***Definice velikosti množiny***

* velikost množiny A: |A|
  + definována jako přirozené číslo n právě tehdy, pokud existuje bijekce f: n –> A
  + viz konstrukce přiroz. čísel pomocí množin
* nekonečné množiny:
  + A je spočetná právě tehdy, pokud existuje bijekce f: N –> A
  + –> spočetné množiny mají stejný počet prvků jako přirozená čísla
  + N (množina přirozených čísel), Z (množina celých čísel), Q (množina racionálních čísel) jsou spočetné množiny
  + R (reální čísla) – není spočetná množina
  + A má mohutnost kontinua právě tehdy, pokud existuje bijekce f: R –> A
* **POSLOUPNOSTI A ŘADY**
* ***Posloupnosti***
* definice: funkce p: N –> R je posloupnost
  + mimo fčního zápisu p(n) se často používá indexová forma pn
  + jedná se o skupinu prvků, v nichž nazáleží na pořadí prvků
* konečné posloupnosti – tj. uspořádané n-tice
* nekoncečné posloupnosti
  + tj. fce na přirozených číslech
  + a0, a1, ..., an je jen jiný zápis f(0), f(1),...f(n)
* induktivní definice nekočné posloupnosti
  + vypíšeme jen 1. člen (event. prvních několik členů)
  + určíme předpis podle něhož dostaneme an s pomocí an-1 (případně an-2 apod.)
* ***Fibonacciho posloupnost***
* 0, 1, 1, 2, 3, 5, 8, 13, 21, 34
* a0 = 0
* a1= 1
* an = an-1 + an-2

**3. Statistika a pravděpodobnost – statistický soubor a jeho charakteristiky, pravděpodobnostní prostor**

**4. Aplikace statistiky na zpracování jazyka**

**5. Statistický jazykový (n-gramový) model – konstrukce, využití**

**6. Vyhodnocování aplikací zpracování jazyka – přesnost, pokrytí, F-míry a zlatý standard, křížová validace**

1. **intenze** neboli smysl pojmu je jeho obsah, to, co se pojmem míní, jeho hlavní či ústřední význam [↑](#footnote-ref-2)