

# Hodnocení kvality sumarizátorů textů

Josef Steinberger<sup>1</sup>, Karel Ježek<sup>1</sup>

<sup>1</sup>Katedra informatiky a výpočetní techniky, FAV, ZČU – Západočeská Univerzita v Plzni,  
Univerzitní 22, 306 14 Plzeň  
{jstein, jezek\_ka}@kiv.zcu.cz

**Abstrakt.** Příspěvek se zabývá možnostmi hodnocení kvality výsledků sumarizátorů textů. Jsou zde popsány jednotlivé třídy metod hodnocení. Větší pozornost je věnována třídě metod založených na podobnosti obsahu. Uvádíme dvě nové metody hodnocení kvality extraktů, které využívají singulární dekompozici. Článek prezentuje výsledky testování tří sumarizátorů hodnocených z hlediska standardní kosinové metody a dvou nových metod založených na singulární dekompozici. Podobnost obsahu extraktu je měřena jak k plnému textu tak i k jeho abstraktu.

**Klíčová slova:** sumarizace, extrakt, abstrakt, singulární dekompozice

## 1 Úvod

Automatická sumarizace textů je vědecká disciplína, která v dnešní době poutá stále větší pozornost. Obrovské množství elektronických informací se musí redukovat, aby se s nimi uživatel mohl efektivněji vypořádat. Existuje mnoho sumarizačních metod. První skupinu tvoří extraktivní metody (*Summarization by Extraction*), jejichž cílem je vybrat nejvýznamnější věty v sumarizovaném textu a výsledkem je tedy extrakt. V dnešní době se začínají vyvíjet také algoritmy, které by umožňovaly automatickou tvorbu abstraktu generováním nových vět (*Summarization by Generation*). Tyto metody jsou však stále málo dokonalé. Pokud však chceme vylepšovat sumarizační metody, musíme nejprve vyvinout způsoby, jak posuzovat jejich kvalitu. Tento příspěvek se zabývá přístupy k hodnocení extraktů. Metody lze rozdělit do dvou skupin. První tvoří metody, které zjišťují, nakolik je automatický systém schopný zachytit hlavní myšlenky zdrojového dokumentu (*content evaluation*). Metody druhé skupiny měří textovou kvalitu výstupu sumarizátoru – hodnotí čtivost, gramatiku a logickou souvislost extraktu (*text quality evaluation*).

Dále je možné rozdělit metody na vnitřní (*intristic*) a vnější (*extrinsic*). První skupina metod vychází z obecných požadavků na abstrakty. Zkoumá tedy informativnost, pokrytí a správnost abstraktů (extraktů). Druhá zkoumá jejich kvalitu vzhledem k určitému cílovému použití.

V aktuálním výzkumu jsou dominantní následující řešení:

1. Srovnání extraktu s referenčním textem (vnitřní hodnocení) – viz kap. 2, 3 a 5
2. Subjektivní hodnocení soudců (vnitřní hodnocení)
3. Aplikačně založené hodnocení (vnější hodnocení) – viz kap. 4

Pro extrakty vět je často měřena podobnost výběru vět extraktu s referenčním extraktem (*co-selection*). Jako hlavní metriky se používají přesnost, úplnost a F-skóre [5]. Tyto metody sebou však přináší také některá omezení. Například lze je použít pouze pro extraktivní sumarizátory. Tuto nevýhodu odstraňují metody založené na podobnosti obsahu (*content-based*), které měří podobnost s referenčním extraktem na úrovni slov [5].

Aplikačně založené metody (*task-based*) měří kvalitu extraktů z hlediska jejich použití pro určitý úkol. Příkladem je kategorizace textů (viz kap. 4).

## 2 Metody založené na společném výběru částí textu

Tyto metody měří kvalitu z hlediska společného výskytu částí textu (většinou vět) v extraktech. Je zde nutné mít k dispozici referenční extrakt neboli extrakt, který se považuje za správný (*gold standard*). Tento extrakt se nejčastěji získává tak, že několik lidí (soudců) vybere věty, které by dle jejich mínění neměly v extraktu chybět. Věty, které jsou vybrány nejvíce soudci se zařadí do referenčního extraktu. V další části uvádíme tři možnosti měření kvality.

### 2.1 Přesnost a úplnost

Koeficienty přesnosti a úplnosti jsou definovány jako:

$$\text{Koeficient přesnosti (Precision):} \quad P(E) = \frac{A}{A + C} \quad (1)$$

$$\text{Koeficient úplnosti (Recall):} \quad R(E) = \frac{A}{A + B} \quad (2)$$

E = vyhodnocovaný extrakt

A = počet shodných vět ve vyhodnocovaném a referenčním extraktu

C = počet vět, které jsou ve vyhodnocovaném extraktu, ale nejsou v referenčním

B = počet vět, které jsou v referenčním extraktu, ale nejsou ve vyhodnocovaném

### 2.2 Kappa

Kappa je statistická míra, která má následující výhody:

- Vylučuje náhodnou shodu, která je definována jako úroveň kvality, která bude dosažena náhodným výběrem vět.
- Umožňuje určení pravděpodobnosti shody mezi soudci.

Koeficient Kappa (K) je definován následovně:

$$K = \frac{P(S) - P(N)}{1 - P(N)}, \quad (3)$$

kde  $P(S)$  je pravděpodobnost shody mezi systémem (resp. soudcem) a referenčním extraktem (resp. jiným soudcem) a  $P(N)$  je pravděpodobnost náhodné shody.

Tedy  $K=0$ , pokud je shoda sumarizačního systému stejná jako při náhodné shodě a  $K=1$ , pokud je hodnocený extrakt shodný s referenčním (perfektní shoda). Pokud je shoda horší, než by se očekávalo při náhodné shodě,  $K$  může být také záporné.

### 2.3 Relativní přínos

Jednotliví soudci nejprve ohodnotí věty (0-10). Ohodnocení automatických extraktů se potom zvyšuje přítomností vět s vysokým ohodnocením a snižuje přítomností redundantních vět [5].

## 3 Metody založené na podobnosti obsahu

Manuální extrakty (vytvořené soudci) však obecně nesdílejí mnoho stejných vět, a proto je problém získat referenční extrakt. Další nevýhodou výše popsaných metod je neschopnost zachytit sémantiku vět. Porovnejme například význam následujících dvou vět:

S1: „Návštěva prezidenta Spojených států amerických v Číně“

S2: „Americký prezident navštívil Čínu“

Do referenčního extraktu bude zařazena pouze jedna z těchto vět. Při hodnocení metodami založenými na společném výběru vět se bude přítomnost druhé věty v extraktu brát jako chyba.

S těmito nepříjemnými vlastnostmi se lze vypořádat použitím metod založených na podobnosti obsahu. Tyto metody počítají podobnost extraktů na nižší úrovni než jenom na úrovni celých vět.

Nabízí se zde tři možnosti:

- počítat podobnost vyhodnocovaného extraktu s referenčním extraktem,
- počítat průměr z podobností vyhodnocovaného extraktu s jednotlivými manuálními extrakty,
- počítat podobnost vyhodnocovaného extraktu s plným dokumentem.

Následují tři běžná měření podobnosti.

### 3.1 Kosinová podobnost

Kosinová podobnost (*Cosine Similarity*) se počítá podle následujícího vzorce:

$$\cos(X, Y) = \frac{\sum x_i * y_i}{\sqrt{\sum (x_i)^2} * \sqrt{\sum (y_i)^2}}, \quad (4)$$

kde  $X$  a  $Y$  jsou reprezentace textu ve vektorovém modelu. U kosinové podobnosti se pro reprezentace textu používají slova nebo lemmata.

### 3.2 Překrytí obsahu

Překrytí obsahu (*Unit Overlap*) lze spočítat následujícím způsobem:

$$\text{overlap}(X, Y) = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|}, \quad (5)$$

kde  $X$  a  $Y$  jsou množinové reprezentace textu.  $|S|$  je mohutnost množiny  $S$ . Zde se jako jednotky reprezentace textu používají množiny slov nebo lemmat.

### 3.3 Nejdelší subsekvence

Nejdelší subsekvenci (*Longest Common Subsequence*) můžeme spočítat následující formulí:

$$\text{lcs}(X, Y) = \frac{\text{length}(X) + \text{length}(Y) - \text{edit}_{dt}(X, Y)}{2}, \quad (6)$$

kde  $\text{length}(X)$  je počet slov řetězce  $X$  a  $\text{edit}_{dt}(X, Y)$  je minimální počet operací vložení a vymazání slov potřebný k transformaci  $X$  na  $Y$ .

## 4 Aplikačně založené metody

Aplikačně založené metody zjišťují kvalitu skrze užití automatických extraktů pro daný praktický úkol. Testovat je možné například zvýšení rychlosti či přesnosti vyhledávání dokumentů, pokud je vyhledávání založené na extraktech místo na plných dokumentech (Korelace relevance). Dalším možným měřením je úspěšnost kategorizace dokumentů do tématických skupin, pokud se indexují extrakty místo původních dokumentů.

### 4.1 Korelace relevance

Korelace relevance (*RC - Relevance correlation*) je nová technika, která umožňuje měřit relativní pokles výkonu získávání informací, pokud se indexují extrakty místo plných dokumentů [5]. Předpokládejme, že máme dotaz  $Q$  a kolekci dokumentů  $D_i$ , vyhledávací systém seřadí dokumenty  $D_i$  podle jejich relevance k dotazu  $Q$ . Potom provedeme substituci plných dokumentů za extrakty  $S_i$  a stejný vyhledávací systém seřadí dokumenty  $S_i$  podle jejich relevance k dotazu  $Q$ . Pokud jsou extrakty dobrou náhradou původních dokumentů, předpokládá se, že pořadí v obou případech budou podobná. Existuje několik metod pro měření podobnosti pořadí (Kendall's tau [8], Spearman's rand correlation [8]). Protože však máme navíc k dispozici (z vyhledávacího systému) relevanci jednotlivých dokumentů k dotazu, můžeme spočítat *RC* následujícím způsobem:

$$RC = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}, \quad (7)$$

kde  $x_i$  je relevance dokumentu  $D_i$  k dotazu  $Q$ ,  $y_i$  je relevance dokumentu  $S_i$  k dotazu  $Q$ .  $\bar{x}$  (resp.  $\bar{y}$ ) je průměrná relevance dokumentů  $D_i$  (resp.  $S_i$ ) k dotazu  $Q$ .

#### 4.2 Kategorizace dokumentů

Tato metoda zjišťuje vhodnost použití extraktů místo plných textů pro pozdější kategorizaci. Pro měření je potřebná zatříděná kolekce dokumentů. Při tomto způsobu testování se ke klasifikaci používá automatický klasifikátor. Z důvodu oddělení chyby klasifikátoru a chyby sumarizátoru je pak nutné použití některých základních hodnot pro porovnání. Výsledné hodnoty klasifikace extraktů jsou proto porovnávány např. s výsledky hodnocení původních dokumentů nebo hodnocení náhodně vybraných vět.

Posledním problémem zůstává míra určující kvalitu extraktu. Obecně se používají koeficienty přesnosti kategorizace  $P$  a úplnosti kategorizace  $R$ :

$$P = \frac{p}{q}, \quad (8)$$

$$R = \frac{p}{r}, \quad (9)$$

kde  $p$  je počet tříd, do kterých je dokument správně zatříděn klasifikátorem,  $q$  je celkový počet tříd, do kterých je dokument klasifikátorem zařazen a  $r$  je počet relevantních tříd, do kterých byl dokument klasifikován při předchozím ručním zatřídění. Potom  $P$  a  $R$  pro celou kolekci je průměrem  $P$  a  $R$  přes všechny dokumenty. Z definice je možné vidět, že oba ukazatele spolu souvisí a zvyšováním jednoho se druhý bude snižovat. Při zařazení dokumentu do co nejvyššího počtu tříd bude vysoká úplnost, při snižování počtu tříd se bude zvedat přesnost. Z toho důvodu se pak používá pro hodnocení klasifikace např. průměr z obou hodnot [4].

## 5 Hodnocení singulární dekompozicí

Singulární dekompozice je proces, který je schopen najít v dokumentu hlavní témata ( $n$ -tice slov) a hodnoty jejich významnosti. Tuto vlastnost využívají naše dvě nová hodnocení, která zjišťují podobnost hlavního tématu (respektive  $n$  hlavních témat) automatického extraktu a referenčního dokumentu. Protože jsou tyto metody založeny na podobnosti obsahu, lze jako referenční dokument použít i manuální abstrakt, který může obsahovat i nově vytvořené věty.

## 5.1 Singulární dekompozice (Singular Value Decomposition)

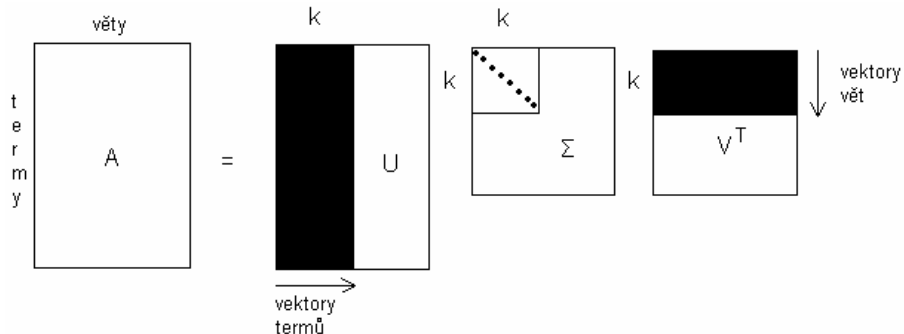
Singulární dekompozice (SVD) je numerický proces, který se hojně používá při redukcí dat. V poslední době byly navrženy algoritmy, které singulární dekompozici řeší klasifikaci nebo vyhledávání dokumentů (latentní sémantické indexování). SVD byla již použita i pro sumarizaci [3]. Na této myšlence jsou založeny i naše hodnotící algoritmy.

Proces začíná vytvořením matice termů proti větám  $A = [A_1, A_2, \dots, A_n]$ , kde každý sloupcový vektor  $A_i$  reprezentuje vektor frekvencí termů ve větě  $i$  v dokomponovaném dokumentu. Pokud dokument obsahuje  $m$  termů a  $n$  vět, získá se matice  $A$  o rozměrech  $m \times n$ . Matice  $A$  je zpravidla řídká, protože normálně se každé slovo v každé větě nevyskytuje. Singulární dekompozice matice  $A$  je potom definována jako:

$$A = U\Sigma V^T, \quad (10)$$

kde  $U = [u_{ij}]$  je  $m \times n$  sloupcově ortonormální matice, jejíž sloupce se nazývají levé singulární vektory,  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$  je  $n \times n$  diagonální matice, jejíž diagonální prvky jsou nezáporná singulární čísla seřazená sestupně a  $V = [v_{ij}]$  je  $n \times n$  ortonormální matice, jejíž sloupce se nazývají pravé singulární vektory (viz obr. 1). Pokud  $r$  je řád matice  $A$ , potom (viz [2])  $\Sigma$  splňuje:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_n = 0. \quad (11)$$



Obr. 1. Singulární dekompozice

Na interpretaci aplikování SVD na matici termů proti větám se můžeme dívat ze dvou pohledů. Z transformačního pohledu SVD zprostředkovává mapování mezi  $m$ -dimenzionálním prostorem vektorů frekvencí termů a  $r$ -dimenzionálním singulárním vektorovým prostorem. Ze sémantického pohledu SVD poskytuje latentní sémantickou strukturu dokumentu reprezentovaného maticí  $A$ . Tato operace vyjadřuje rozklad originálního dokumentu do  $r$  lineárně nezávislých bázevých vektorů. Každý term i věta dokumentu jsou indexovány těmito bázevými vektory. Unikátní vlastností singulárního rozkladu je schopnost zachytit a modelovat vnitřní vztahy mezi termy tak, že může sémanticky shlukovat termy a věty. Dále, jak je demonstrováno v [2],

pokud se v dokumentu často vyskytuje určitá kombinace slov, pak bude tato kombinace zachycena a reprezentována jedním ze singulárních vektorů. Velikost odpovídajícího singulárního čísla indikuje významnost kombinace v dokumentu. Každá věta obsahující tuto kombinaci slov bude promítnuta blízko odpovídajícího singulárního vektoru a věta, která nejlépe reprezentuje tuto kombinaci bude mít největší hodnotu v tomto vektoru. Každá kombinace slov popisuje určité téma dokumentu. Lze tedy na základě předchozích faktů říci, že každý singulární vektor reprezentuje určité téma dokumentu a velikost korespondujícího singulárního čísla reprezentuje významnost tohoto tématu [3].

Na základě předchozí diskuse jsme navrhli metodu hodnotící kvalitu extraktů. Tato metoda využívá singulární rozklad matice termů proti větám, konkrétně matici  $U$ , která popisuje míru významnosti termů v hlavních tématech (singulárních vektorech) dokumentu. Při hodnocení měříme podobnost mezi maticí  $U$  získanou singulárním rozkladem originálního dokumentu (nebo referenčního extraktu nebo abstraktu) a maticí  $U$  získanou singulárním rozkladem hodnoceného extraktu. Pro zjištění této podobnosti jsme navrhli následující dvě měření.

## 5.2 Podobnost hlavního tématu

Tento způsob měření porovnává první levé singulární vektory získané singulárním rozkladem referenčního dokumentu (plný text, referenční extrakt nebo abstrakt) a extraktu. Tyto vektory odpovídají hlavnímu tématu dokumentu. Čím více se tedy podobá hlavní téma extraktu hlavnímu tématu referenčního dokumentu, tím je extrakt kvalitnější. Podobnost měříme úhlem mezi vektory, které jsou normalizované, takže můžeme použít následující vzorec:

$$\cos \varphi = \sum_{i=1}^n ue_i \cdot ur_i, \quad (12)$$

kde  $ur$  je první levý singulární vektor rozkladu referenčního dokumentu,  $ue$  je první levý singulární vektor extraktu (hodnoty odpovídající termům jsou uspořádány podle referenčního dokumentu a na místě chybějících termů jsou nuly),  $n$  je počet různých termů referenčního dokumentu.

## 5.3 Podobnost $n$ nejvýznamnějších témat

Tento způsob hodnotí podobnost z pohledu  $n$  hlavních témat porovnávaných dokumentů. Nejprve vytvoříme singulární rozklady referenčního dokumentu a hodnoceného extraktu. Potom pro oba dokumenty vynásobíme matice  $U$  a  $\Sigma$ . Získáme tím matice  $S_e$  (pro extrakt) a  $S_r$  (pro referenční dokument):

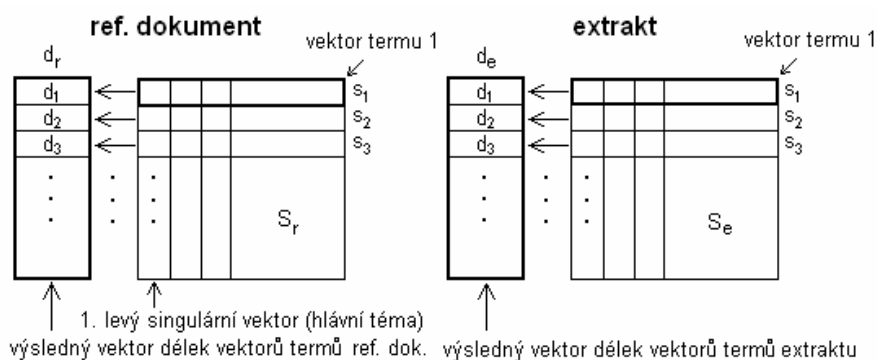
$$S_e = U_e \cdot \Sigma_e, \quad (13)$$

$$S_r = U_r \cdot \Sigma_r. \quad (14)$$

Důvodem tohoto násobení je zvýhodnění hodnot v matici  $U$ , které patří k nejvyšším singulárním číslům (nevýznamnějším tématům). Pro každý vektor termu (viz obr. 2) pak spočítáme jeho délku (pro extrakt i referenční dokument):

$$d_k = \sqrt{\sum_{i=1}^n s_{k,i}^2}, \quad (15)$$

kde  $d_k$  je délka vektoru  $k$ -tého termu,  $n$  je počet témat. V našich experimentech jsme zvolili konstantní počet tří nevýznamnějších témat.



**Obr. 2.** Vytvoření výsledných vektorů ref. dokumentu a extraktu

Z délek vektorů termů sestavíme výsledný vektor délek termů v latentním prostoru vzniklém singulární dekompozicí (viz obr. 2). Získáme tím tedy dva výsledné vektory – jeden pro extrakt ( $d_e$ ) a druhý pro referenční dokument ( $d_r$ ). Tyto vektory potom znormalizujeme. Pro změření jejich podobnosti použijeme opět kosinovou míru – analogicky s (12):

$$\cos \varphi = \sum_{i=1}^n d_{e_i} \cdot d_{r_i}. \quad (16)$$

Tato metoda má následující výhodu oproti předchozí. Pokud budeme uvažovat originální dokument jako referenční a ten bude obsahovat dvě přibližně stejně důležitá témata (odpovídající singulární čísla budou mít přibližně stejnou hodnotu), pak se může stát, že v extraktu převládne druhé nevýznamnější téma nad prvním a hlavní téma bude vyhodnoceno jako velmi rozdílné. Tuto nevýhodu odstraníme, pokud hodnotíme podle více témat.

## 6 Testování

### 6.1 Testovací kolekce

K testování nových metod jsme použili kolekci Computation and Language Collection (cmp-lg) [7], která byla vytvořena korporací MITRE a univerzitou v Edinburghu. Tato kolekce formátu xml slouží jako zdroj pro testování v oblasti



dolování znalostí a sumarizace textů. Vyskytují se zde vědecké dokumenty dlouhé průměrně 169 vět a také jsou zde jejich abstrakty.

**Tabulka 1.** Detaily o testovací kolekci cmp-lg

Počet dokumentů	178
Minimální počet vět v dokumentu	45
Maximální počet vět v dokumentu	750
Průměrný počet vět v dokumentu	169
Průměrný počet slov v dokumentu	4504
Průměrný počet významových slov v dokumentu	1653
Průměrný počet různých významových slov v dokumentu	529

## 6.2 Výsledky

Úkolem testování bylo porovnání nových hodnocení, založených na singulární dekompozici, se standardní kosinovou podobností. Porovnávali jsme extrakty tří sumarizátorů. Prvním je náš sumarizátor, založený na myšlence sumarizování singulární dekompozicí [6]. Druhým je náhodný sumarizátor (RANDOM), který vybírá náhodné věty dokumentu. Ten slouží ke stanovení spodní hranice úspěšnosti. Třetím je sumarizátor se zabudovanou poziční heuristikou (LEAD) [5]. Ten preferuje věty na začátku dokumentu. Z výsledků lze tedy zkoumat, zda a nakolik se vyskytují důležité informace na začátku dokumentu. Nejprve jsme jako referenční dokument vzali plný text. Zkoumali jsme tedy podobnost plného textu a jeho extraktu. Potom jsme porovnávali abstrakt, který byl v kolekci připojen ke každému článku, a automatický extrakt. Sumarizační poměr byl stanoven na 3%, 5% a 10%. Jako sumarizační jednotku jsme nejprve brali větu, ale toto nastavení penalizovalo náhodné a poziční extrakty, které mohou vybrat i velmi krátké věty a celé extrakty tedy mohou být podstatně kratší. Proto bylo nakonec jako sumarizační jednotka nastaveno slovo a všechny extrakty byly přibližně stejně dlouhé. Důležitým faktorem kvality singulárního rozkladu je kvalita lematizačního slovníku. Při našem testování jsme použili lematizační slovník (SMART's English Stoplist), jehož autory jsou M.W. Berry a S. Dumais [1]. Hodnoty v následujících tabulkách udávají kosinus odchylky vektorů.

**Tabulka 2.** Hodnocení kvality 3% extraktů (ref. dokument je plný text)

Hodnocení	Sumarizátor		
	SVD	RANDOM	LEAD
Kosinová podobnost	<b>0.651</b>	0.521	0.426
Podobnost hlavního tématu	<b>0.694</b>	0.390	0.371
Podobnost třech hlavních témat	<b>0.650</b>	0.420	0.378

**Tabulka 3.** Hodnocení kvality 5% extraktů (ref. dokument je plný text)

Hodnocení	Sumarizátor		
	SVD	RANDOM	LEAD
Kosinová podobnost	<b>0.712</b>	0.603	0.504
Podobnost hlavního tématu	<b>0.767</b>	0.453	0.449
Podobnost třech hlavních témat	<b>0.653</b>	0.486	0.449

**Tabulka 4.** Hodnocení kvality 10% extraktů (ref. dokument je plný text)

Hodnocení	Sumarizátor		
	SVD	RANDOM	LEAD
Kosinová podobnost	<b>0.788</b>	0.733	0.617
Podobnost hlavního tématu	<b>0.858</b>	0.608	0.565
Podobnost třech hlavních témat	<b>0.722</b>	0.600	0.556

V předchozích tabulkách je vidět, že u kosinové podobnosti není příliš velký rozdíl mezi propracovanou sumarizační technikou (SVD) a náhodným výběrem vět. U obou hodnocení singulární dekompozicí je rozdíl podstatně větší. Poziční sumarizátor byl vyhodnocen jako horší než náhodný.

**Tabulka 5.** Hodnocení kvality 3% extraktů (ref. dokument je abstrakt)

Hodnocení	Sumarizátor		
	SVD	RANDOM	LEAD
Kosinová podobnost	0.543	0.240	<b>0.551</b>
Podobnost hlavního tématu	<b>0.325</b>	0.095	0.295
Podobnost třech hlavních témat	<b>0.308</b>	0.100	0.284

**Tabulka 6.** Hodnocení kvality 5% extraktů (ref. dokument je abstrakt)

Hodnocení	Sumarizátor		
	SVD	RANDOM	LEAD
Kosinová podobnost	0.572	0.271	<b>0.604</b>
Podobnost hlavního tématu	<b>0.353</b>	0.110	0.341
Podobnost třech hlavních témat	0.319	0.119	<b>0.335</b>

**Tabulka 7.** Hodnocení kvality 10% extraktů (ref. dokument je abstrakt)

Hodnocení	Sumarizátor		
	SVD	RANDOM	LEAD
Kosinová podobnost	0.599	0.308	<b>0.669</b>
Podobnost hlavního tématu	0.387	0.149	<b>0.403</b>
Podobnost třech hlavních témat	0.353	0.147	<b>0.380</b>

Dále jsme zjišťovali podobnost extraktů s abstrakty. U 3% extraktů byl vyhodnocen SVD sumarizátor jako nejlepší, ale u delších extraktů (5% a 10%) jej již převýšil poziční sumarizátor. Toto zdánlivě překvapující zjištění potvrdilo, že na začátku článků se vyskytují významné informace. Je to vlastnost, které využívají např. heuristické sumarizátory (poziční heuristika – vyšší ohodnocení vět v úvodcích článků).

## 7 Závěr

Tento příspěvek představil moderní trendy při hodnocení kvality sumarizátorů. Popsali jsme také novou hodnotící metodu založenou na singulární dekompozici. Testování prokázalo, že nová metoda je schopna rozlišit kvalitní extrakty od náhodných výběrů vět lépe než dnes používaná kosinová metoda. Dále jsme zaznamenali, že SVD je velice citlivá na kvalitu stoplistu a lematizační proces. V dalším výzkumu plánujeme zkvalitnění výstupu našeho SVD sumarizátoru. Budou zkoumány následující možnosti – rezoluce anafor, redukce a kombinace vět a vyloučení podobných vět. Cílem je vyvinout inteligentní sumarizační systém, který bude schopen prezentovat koherentní extrakty, které by správně vystihovaly původní dokument. Je zřejmé, že bez kvalitních hodnotících metod se neobejdeme.

*Příspěvek vznikl za částečné podpory výzkumného záměru MSM 235200005 a ME494.*

## Reference

1. Berry M.W., Dumais S. <http://www.cs.utk.edu/~lsi/>.
2. Berry M.W., Dumais S.T., O'Brien G.W. Using Linear Algebra for Intelligent Information Retrieval. *SIAM Review*. 1995.
3. Gong Y., Liu X. Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. *Sborník 24. konference ACM SIGIR 2001*. New Orleans, USA 2001.
4. Hynek J., Ježek K. Practical Approach to Automatic Text Summarization. *Sborník 7. konference ELPUB '03*. Guimaraes, Portugalsko 2003.

5. Radev R., Teufel S., Saggion H., Lam W., Blitzer J., Qi H., Celebi A., Liu D., Drabek E. Evaluation Challenges in Large-scale Document Summarization. *Sbornik 41. konference ACL 2003*. Sapporo, Japonsko 2003.
6. Steinberger J., Ježek K. Text Summarization and Singular Value Decomposition. *Sbornik 3. konference ADVIS '04*. Izmir, Turecko 2004.
7. TIPSTER Text Summarization Evaluation Conference (SUMMAC). [http://www-nlpir.nist.gov/related\\_projects/tipster\\_summac/cmp\\_lg.html](http://www-nlpir.nist.gov/related_projects/tipster_summac/cmp_lg.html)
8. Siegel S., Castellan N. J. Jr. Nonparametric Statistics for the Behavioral Sciences. Berkeley, CA: McGraw-Hill, 2nd edn., 1988

**Annotation:**

*Quality Evaluation of Text Summarizers*

This paper describes possibilities of summary quality evaluation. Firstly, we mention the taxonomy of evaluation approaches. Further, we pay close attention to content-based methods. Then, two new evaluation methods that use singular value decomposition are proposed. Finally, we present evaluation results for three different summarizers from the angle of standard cosine content-based method and the two new evaluation methods based on the singular value decomposition. In the tests the content similarity between an original document and its summary and between the summary and its corresponding abstract was compared.