Pokroky matematiky, fyziky a astronomie

Jan Brandts; Michal Křížek
Lineární algebra ukrytá v internetovém vyhledávači Google

Pokroky matematiky, fyziky a astronomie, Vol. 52 (2007), No. 3, 195–204

Persistent URL: `http://dml.cz/dmlcz/141358`

Terms of use:

© Jednota českých matematiků a fyziků, 2007

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* `http://dml.cz`

Lineární algebra ukrytá v internetovém vyhledávači Google

Jan Brandts, Amsterdam, a Michal Křížek, Praha

1. Krátce z historie

Během posledních deseti let Google upevnil své postavení nejen jako nejpoužívanější vyhledávací software na internetu, ale též jako velice úspěšná společnost, která vyvinula i mnoho dalších užitečných programů (např. Google Earth, Google Video, Google Scholar, Gmail, Google Docs & Spreadsheets, Google Book Search viz [8–11]). Slovo google přešlo i do jazyka. Anglicky „to google“ znamená najít něco na internetu pomocí vyhledávače Google. I česky se začalo celkem běžně používat sloveso „vygooglovat“. Samotné slovo google vzniklo ze slova „googol“, které bylo poprvé vysloveno v roce 1920 devítiletým chlapcem Miltonem Sirottou (viz [6]) pro nesmírně obrovské číslo 10^{100}. Toto číslo je mnohem větší, než je počet všech atomů v pozorovatelné části vesmíru, ale zase menší než počet všech možností, jak může stát za sebou 70 lidí ve frontě.

Internetový vyhledávač Google vyvinuli v roce 1998 Ph. D. studenti Larry Page a Sergey Brin (viz obr. 1) ze Stanfordské univerzity. Jeho architektura je popsána v [1]. Tento vyhledávač provádí analýzu webovských stránek a odkazů na ně k nalezení potřebné informace. Získávání informací od zrodu celosvětové počítačové sítě (World Wide Web) v roce 1989 už nikdy nebude stejné jako dříve. Místo hledání relevantní odpovědi na daný dotaz na již poněkud zastarávajících médiích, jako jsou např. encyklopedie, kartotéky, diskety, mikrofilmy, cédéčka, stačí v současnosti napsat na klávesnici počítače jeden či několik hledaných termínů (klíčových slov) a takřka okamžitě dostanete nějaké odpovědi anebo zprávu, že vyhledávací zařízení hledaný termín nenalezlo. Této obrovské rychlosti se dosahuje především díky předzpracování prohledávaných dat (jejich organizací do speciálních datových struktur), paralelizaci prohledávání stránek, použitím efektivních numerických algoritmů a vysokou výkonností soudobých počítačů.

Celosvětová počítačová síť tak zřejmě poskytuje obrovské možnosti, s nimiž jsme se nikdy dříve nesetkali. Je nesmírně rozsáhlá a je to vlastně největší technické zařízení na světě. Má přibližně 10 miliard webovských stránek o průměrné velikosti 500 KB. Při

Dr. Jan Brandts (1968), Korteweg-de Vries Institute, Faculty of Science, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, Nizozemí, e-mail: brandts@science.uva.nl

Prof. RNDr. Michal Křížek, DrSc. (1952), Matematický ústav Akademie věd ČR, v. v. i., Žitná 25, 115 67 Praha 1, e-mail: krizek@math.cas.cz

Obr. 1. Sergey Brin (vlevo) a Larry Page (vpravo).

zadání dotazu se musí Google s tímto obrovským množstvím informace nějak vypořádat. Přitom cca 40 procent stránek se mění každý týden a zhruba 23 procent dokonce každý den. Web je tedy velice dynamický. Navíc se organizuje „sám" v tom smyslu, že zde nejsou žádní trénovaní specialisté jako knihovníci, kteří by nesli odpovědnost za obsah stránek. Na druhé straně jsou zde také hackeři a tvůrci spamů, kteří zneužívají svobody internetu a záměrně se pokoušejí způsobit na webu chaos nebo dokonce usilují o jeho destrukci.

Nejdůležitější rozdíl mezi webem a jinými informačními systémy spočívá v tom, že řada webovských stránek má množství odkazů na jiné stránky. Web lze tedy chápat jako orientovaný graf, což přináší novou kvalitu do této informační struktury. Poměrně dlouho trvalo, než si Jon Kleinberg[1]) z firmy IBM uvědomil (též v roce 1998), že ke studiu takové struktury lze použít matematické metody. A právě tyto nové možnosti byly plně zúročeny v internetovém vyhledávači Google a také v méně známém systému HITS (Hypertext Induced Topic Search), jejž vyvinul sám Kleinberg a který se používá ve vyhledávači Teoma.

Při psaní tohoto článku jsme se inspirovali zejména nedávno vydanou knihou [3], kterou vřele doporučujeme k dalšímu studiu.

2. Oceňování webovských stránek

Úspěch Googlu spočívá v naprosto jedinečném a velice elegantním způsobu oceňování webovských stránek. Každá stránka má přiřazeno jisté číslo (váhu z intervalu [0,1]), které charakterizuje její důležitost a kterému budeme říkat stránkové ocenění (angl. PageRank[2])). Jestliže má internet v současnosti přibližně 10 miliard stránek, pak jejich ocenění je dáno vektorem stejné délky. Tento vektor lze dobře definovat a určit pomocí prostředků numerické lineární algebry. S vektorem stránkového ocenění

[1]) V roce 2006 získal J. Kleinberg Nevanlinovu cenu (částečně i za práce související s Googlem).

[2]) Googlovská PageRank metoda se používá i k vyjádření prestiže vědeckých časopisů — viz [7, s. 34].

a příslušnými maticemi, které jej definují, musí být snadné manipulovat. Rovněž výpočetní rychlost vyžaduje zvláštní pozornost. Google tak poskytuje zajímavé problémy nejen z informatiky, ale i z celé řady matematických odvětví, jako je např. teorie grafů, lineární algebra, numerická matematika, výpočetní složitost, zpracování dat, matematická lingvistika, umělá inteligence, matematické modelování, Markovovské řetězce a statistika. Bez podpory těchto disciplín by byla existence vyhledávače Google nemyslitelná.

2.1. Základní model

Nyní popíšeme jednoduché algebraické vztahy, které definují vektor stránkového ocenění. Opírají se o jisté heuristické postupy, které se kombinují s matematickými metodami. To může být někdy protichůdné, a proto musí být učiněn jistý kompromis. Uvidíme, že vystačíme pouze se základními pojmy vektorového a maticového počtu.

Velmi spolehlivé fungování Googlu je založeno na několika heuristických předpokladech a tvrzeních, kterým budeme říkat teze.

Teze 1. *Ocenění webovské stránky je vysoké, jestliže na ni odkazují vysoce oceněné stránky a navíc jestliže tyto stránky zároveň neodkazují na mnoho dalších stránek.*

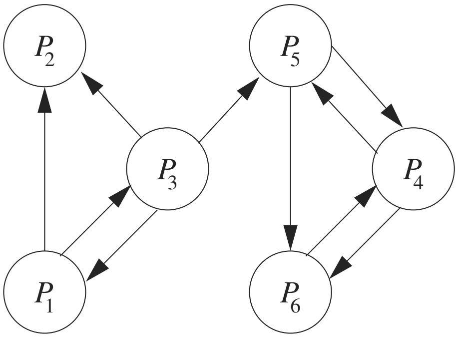
Tato heuristika odpovídá zkušenosti, že webovská stránka je důležitá, jestliže hodně lidí cítí potřebu na ni odkazovat ze svých vlastních stránek. Webovský odkaz může tedy být interpretován jako jakési doporučení. Jistě je lepší, když takové doporučení pochází od nějaké významné autority, a je ještě lepší, když seznam odkazů této autority není příliš velký, ale omezuje se jen na jejich malý počet.

Zaveďme si nyní několik jednoduchých označení, abychom mohli tezi 1 zformulovat poněkud více matematicky. Předpokládejme, že n je celkový počet webovských stránek a že $P_1, \ldots, P_n$ jsou samotné webovské stránky, tj. vrcholy v orientovaném webovském grafu. Nechť $\mathcal{B}_j$ je množina webovských stránek P_i, které odkazují na stránku P_j, a nechť $|P_i|$ je celkový počet odkazů směřujících z P_i. Pak můžeme ocenění stránky P_j „definovat“ takto:

$$r(P_j) = \sum_{P_i \in \mathcal{B}_j} \frac{1}{|P_i|} r(P_i). \tag{1}$$

Ocenění stránky P_j je tedy vážený součet ocenění stránek, které odkazují na P_j. I když se zdá, že je tato definice kruhem, příklad z obrázku 2, který je převzat z knížky [3], ukazuje, že takový vektor čísel $\big(r(P_1), \ldots, r(P_n)\big)$ splňujících rovnici (1) může existovat. Pro jeho výpočet je celkem přirozené použít známou metodu postupných aproximací.

Nechť je dán nějaký počáteční vektor délky n. Pomocí vztahu (1) budeme postupně vypočítávat nové hodnoty $r(P_j)$ a celý proces budeme neustále opakovat. Přirozenou otázkou je, zda takto definovaný iterační proces bude konvergovat. Pokud označíme



Obr. 2. Příklad orientovaného grafu odpovídajícího stránkovému ocenění, které splňuje rovnici (1). Matice reprezentující tento graf je dána vztahem (4) a příslušný vektor stránkového ocenění vztahem (6).

vektor stránkového ocenění $\pi \in \mathbb{R}^n$ a matici H s prvky h_{ij} pro $1 \leqq i, j \leqq n$ jako

$$\pi^\top = \big(r(P_1), \dots, r(P_n)\big) \quad \text{a} \quad h_{ij} = \begin{cases} \dfrac{1}{|P_i|} & \text{pro } P_i \in \mathcal{B}_j, \\ 0 & \text{pro } P_i \notin \mathcal{B}_j, \end{cases} \tag{2}$$

pak vztah (1) lze formulovat ve vektorovém tvaru takto: Nalézt $\pi \in \mathbb{R}^n$, pro něž

$$\pi^\top = \pi^\top H. \tag{3}$$

Tento vztah nám říká, že $\pi^\top$ je levý vlastní vektor matice H odpovídající vlastnímu číslu 1.

Jako příklad uvažujme opět orientovaný graf, jenž je na obrázku 2 a jenž může být reprezentován maticí souvislostí H definovanou vztahem

$$H = \begin{bmatrix} 0 & \frac12 & \frac12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac13 & \frac13 & 0 & 0 & \frac13 & 0 \\ 0 & 0 & 0 & 0 & \frac12 & \frac12 \\ 0 & 0 & 0 & \frac12 & 0 & \frac12 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{4}$$

První řádek H vyjadřuje, že stránka P_1 se odkazuje na stránky P_2 a P_3. Druhý řádek H obsahuje samé nuly, což znamená, že P_2 se neodkazuje na žádné jiné stránky a je to tzv. koncový uzel, atp. Výpočtem se snadno můžeme přesvědčit, že vlastní číslo[3]) matice H je 1 a že odpovídající vlastní vektor je vektor stránkového ocenění odpovídající situaci z obrázku 2.

Výše navrhovaný iterační proces lze nyní přeformulovat takto:

$$\pi_{k+1}^\top = \pi_k^\top H \quad \text{pro} \quad k = 0, 1, 2, \dots, \tag{5}$$

kde

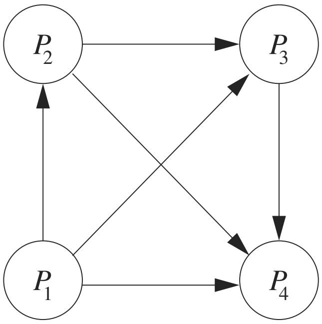
$$\pi_0^\top = \frac{1}{n} e^\top \quad \text{a} \quad e^\top = (1, \dots, 1).$$

[3]) Všechna vlastní čísla matice (4) jsou reálná: $(-\frac12, -\frac12, -\frac25, 0, \frac25, 1)$.

Jde vlastně o standardní mocninnou metodu (viz [2]) pro výpočet vlastního vektoru odpovídajícího známému vlastnímu číslu 1. Pokud použijeme tento iterační proces pro matici H ze vztahu (4), pak lze snadno numericky ověřit, že

$$\pi^\top = \lim_{k\to\infty} \pi_k^\top = \left(0, 0, 0, \tfrac{4}{9}, \tfrac{2}{9}, \tfrac{3}{9}\right). \tag{6}$$

Není ovšem jasné, jestli každá webovská matice souvislostí H má skutečně vlastní číslo rovno jedné, a pokud tomu tak je, zda odpovídající podprostor levých vlastních vektorů je jednorozměrný. Tento poslední požadavek je nutný k tomu, abychom měli jednoznačné pořadí ocenění jednotlivých webovských stránek. Například matice H odpovídající orientovanému grafu z obrázku 3 má pouze nulová vlastní čísla.



Obr. 3. Příklad orientovaného grafu, pro který nelze definovat vektor stránkového ocenění.

Kromě toho konvergence iteračního procesu (5) také není zcela zřejmá. Existuje tedy celá řada problémů, které je třeba ještě vyřešit. Např. není automaticky zaručeno, že

- vždy existuje jediný stacionární vektor π splňující (3),
- iterace (5) k tomuto řešení π konvergují,
- rychlost konvergence bude dostatečná, když $n \approx 10^{10}$,
- se nám podaří nějakým způsobem uložit matici H řádu $n \approx 10^{10}$,
- stacionární vektor $\pi^\top$ od sebe rozliší vzájemně všechny stránky.

Zejména poslední požadavek se nezdá být zcela v pořádku. Kdyby totiž koncové webovské stránky měly ocenění nula, pak bychom nedostali žádnou informaci pro rozhodování. Zbytečně bychom tak promrhali velké množství výpočetního času. Z tohoto důvodu navrhovatelé Googlu požadovali, aby stacionární vektor byl jednoznačně určen a jen s kladnými vzájemně od sebe různými prvky.

2.2. Googlovská rovnice pro stránkové ocenění

Díky lineární algebře a zejména Perronově-Frobeniově teorii (viz [3, s. 167]) víme, že pokud bude webovská matice souvislostí stochastická a kladná (viz definice 1 a 2), potom problém výpočtu vlastního vektoru lze realizovat. Proto Page a Brin navrhli dvě změny původního modelu. První změna se týkala odstranění vlivu nulových řádků v matici H, které jsou způsobeny koncovými webovskými stránkami, které se na nic neodkazují. Zatímco HTML dokumenty mívají obvykle řady odkazů, PDF soubory, obrázky ve formátu JPG a podobné dokumenty se dále na nic neodkazují. Odhaduje se, že přibližně 80 procent webovských stránek je právě tohoto typu. Bylo tedy nezbytné tento rozsáhlý problém vyřešit.

Definice 1. Matice je (*řádkově*) *stochastická*, jestliže jsou všechny její prvky nezáporné a součet prvků v každém jejím řádku je roven 1.

Matici H lze snadno přeměnit na (řádkově) stochastickou, jestliže každý její nulový řádek nahradíme řádkem stejných hodnot, jejichž součet je jedna. To vede na matici S definovanou vztahem

$$S = H + \frac{1}{n}\, ae^{\top}, \text{ kde } a_i = \begin{cases} 1, & \text{jestliže } P_i \text{ je koncová stránka,} \\ 0, & \text{jestliže } P_i \text{ není koncová stránka,} \end{cases} \tag{7}$$

kde $a = (a_1, \dots, a_n)^{\top}$. Tato úprava odpovídá následujícímu modelu.

Teze 2. *Pokud se dostaneme na koncovou webovskou stránku, odskočíme z ní zcela náhodně na jinou stránku.*

Matice S má díky této úpravě vlastní číslo 1. Z definice násobení matice vektorem totiž vidíme, že Se je vektor řádkových součtů matice S, a proto

$$Se = e.$$

Tudíž e je pravý vlastní vektor matice S odpovídající vlastnímu číslu 1. Níže (viz (11)) ukážeme, že S má i levý vlastní vektor odpovídající stejnému vlastnímu číslu.

Nechť $\|w\|_\infty = \max_i |w_i|$ je Čebyševova norma vektoru $w = (w_1, \dots, w_n) \in \mathbb{R}^n$. Potom platí

$$\|Sw\|_\infty \leqq \|w\|_\infty \qquad \text{pro všechna } w \in \mathbb{R}^n,$$

a tedy speciálně i pro každý vlastní vektor. Proto neexistují žádná vlastní čísla matice S, která by byla větší v absolutní hodnotě než 1.

Označme-li I jednotkovou matici, pak

$$0 = \det(S - I) = \det(S^{\top} - I^{\top}) = \det(S^{\top} - I). \tag{10}$$

Odtud plyne, že $S^{\top}$ má také vlastní číslo 1 a jemu odpovídající pravý vlastní vektor označíme opět π, tj.

$$S^{\top}\pi = \pi.$$

Je-li tedy S stochastická, existuje vektor stránkového ocenění π splňující ekvivalentní rovnici

$$\pi^{\top} = \pi^{\top} S. \tag{11}$$

Povšimněte si, že sama jednotková matice I je stochastická a má n-rozměrný prostor vlastních vektorů. Sama stochastičnost matice tedy nestačí k tomu, aby stránkové ocenění bylo jednoznačné. Pokud ale bude navíc S kladná, stránkové ocenění už bude jednoznačné, jak ukážeme později.

Definice 2. Matice se nazývá *kladná*, jestliže všechny její prvky jsou kladné.

Následující poněkud zjednodušená konstrukce tedy bere v úvahu kladnost tzv. googlovské matice.[4]) Označme

$$T = \frac{1}{n}\, ee^\top \tag{12}$$

a definujme *googlovskou matici* G jako netriviální konvexní kombinaci matic S a T, tj. pro daný parametr $\alpha \in (0,1)$ položíme

$$G = \alpha S + (1-\alpha)T. \tag{13}$$

Snadno nahlédneme, že konvexní kombinace stochastických matic je opět stochastická, a tedy i G bude stochastická. Zřejmě G bude také kladná. Podobně jako pro první přizpůsobení googlovské matice existuje další heuristický postup, který motivuje i druhé přizpůsobení.

Teze 3. *Uživatel internetu nepoužívá jen strukturu odkazů, aby se dostal ke svému cíli. Občas do navigačního okna napíše též novou adresu.*

V práci [3] se podle této teze nazývá T *teleportační matice*. Vyjadřuje vlastně zcela náhodný pohyb z jedné webovské stránky na druhou, bez ohledu na strukturu odkazů. Parametr α nejenže určuje množství teleportace, ale hraje také důležitou roli v několika dalších aspektech při řešení *googlovské rovnice stránkového ocenění*:

$$\boxed{\pi^\top = \pi^\top\big(\alpha S + (1-\alpha)T\big).} \tag{14}$$

Abychom ukázali, že G má jednorozměrný podprostor levých vlastních vektorů odpovídajících vlastnímu číslu 1, označíme vektor absolutních hodnot prvků vektoru $v \in \mathbb{R}^n$ symbolem v_+. Pak pro každé $x \in \mathbb{R}^n$, pro něž $-x_+ \neq x \neq x_+$, kladnost matice G společně se vztahem $Ge = e$ dávají, že

$$(x^\top G)_+ e < (x_+{}^\top G)e = x_+^\top e.$$

Odtud vyplývá, že pokud $x^\top = x^\top G$, pak buď $x = x_+$, nebo $x = -x_+$. Nyní již snadno nahlédneme, že podprostor vektorů s touto vlastností je nejvýše jednorozměrný. Jestliže je navíc alespoň jedna složka vektoru x nenulová, pak díky kladnosti G a vztahu $x^\top = x^\top G$ jsou všechny složky nenulové. To dokazuje existenci jediného stránkového ocenění, které se skládá jen z kladných čísel.

3. Výpočet vektoru stránkového ocenění

Kombinace matematických metod a heuristických postupů z předchozí kapitoly vedla k vytvoření modelu s jednoznačným oceněním webovských stránek. Matematicky to

[4]) Ve skutečnosti se místo tohoto předpokladu uvažuje její nerozložitelnost (angl. irreducibility) a aperiodičnost (viz [3]).

lze vyjádřit tak, že googlovská matice G definovaná vztahem (13) má jednorozměrný podprostor levých vlastních vektorů příslušejících vlastnímu číslu 1, který obsahuje vektor s kladnými složkami, jejichž součet je jedna. Tento vektor označme π.

Nyní se vrátíme ke stanovení aproximace vektoru π. Podobně jako ve vztahu (3) vidíme, že nejjednodušší cestou pro její získání je použít následující iterační proces pro výpočet pevného bodu:

$$\pi_{k+1}^\top = \pi_k^\top G \quad \text{pro } k = 0, 1, 2, \ldots, \qquad \text{kde } \pi_0^\top = \frac{1}{n}\, e^\top. \tag{15}$$

I když je googlovská matice plná, její násobení vektorem je relativně levná operace. Přitom se využívá vztahů (7), (12), (13) a toho, že matice H charakterizující web je řídká a že teleportační matice T má hodnost 1.

Protože matice G má obrovský řád $n \approx 10^{10}$, není možné řešit přímo ekvivalentní soustavu lineárních algebraických rovnic, která vznikne ze (14) přidáním rovnice $\pi^\top e = 1$ pro kalibraci vlastního vektoru. Jak již bylo řečeno v kapitole 2, iterační metoda (5), a tedy i (15) není nic jiného než mocninná metoda pro výpočet vlastních hodnot matice. Přitom je známo, že (viz [2])

$$\|\pi - \pi_k\| = \mathcal{O}\biggl(\left|\frac{\lambda_2}{\lambda_1}\right|^k\biggr), \tag{16}$$

kde λ_1 je v absolutní hodnotě největší vlastní číslo matice G a λ_2 je její druhé největší vlastní číslo. V kapitole 2 jsme již ukázali, že $\lambda_1 = 1$ pro G. Bez důkazu (viz [3, s. 46]) poznamenejme, že

$$\lambda_2 = \alpha\mu_2, \tag{17}$$

kde μ_2 je v absolutní hodnotě druhé největší vlastní číslo matice S. Jelikož je S velká stochastická matice, je velká pravděpodobnost, že také $\mu_2 = 1$, což dává rychlost konvergence mocninné metody (15) řádu $\mathcal{O}(\alpha)$. Parametr α z googlovské rovnice (14) pro výpočet stránkového ocenění tak určuje, nejen jak mnoho je teleportace zahrnuta v tomto modelu, ale také řídí rychlost konvergence při výpočtu iterací (15). Je známo, že Google používá přibližně padesát iterací mocninné metody, aby dostal spolehlivě dvě až tři platné cifry při výpočtu vektoru stránkového ocenění. Skutečná hodnota teleportačního parametru α, kterou Google používá, je kolem 0.85.

Vraťme se nyní na chvíli k příkladu z obrázku 2. Pro 15 iteračních kroků mocninné metody s hodnotou parametru $\alpha = 0.85$ a počátečním vektorem $\pi_0^\top = \frac{1}{6}e^\top$ dostaneme

$$\pi_{15}^\top = (0.0517,\ 0.0737,\ 0.0574,\ 0.3487,\ 0.1999,\ 0.2686). \tag{18}$$

Dvě změny v matici H tedy způsobily, že vektor stránkového ocenění neobsahuje nulové prvky (srov. (6)). Navíc se jeho složky ve svých hodnotách v tomto případě vzájemně odlišují.

Algoritmus (15) je relativně snadno proveditelný, i když existují alternativní numerické metody (viz např. [3], [4] a [5]). Vyžaduje minimální nároky na počítačovou paměť a snadno se naprogramuje, zatímco jiné metody mohou mít s velikostí paměti potíže. I přesto se googlovská matice G nevejde do paměti jednoho počítače, ale je uložena ve velkém množství počítačů.

V současnosti Google potřebuje zhruba 3 dny pro stanovení nového vektoru stránkového ocenění. Výpočet se provádí jednou měsíčně, což je poměrně vhodná doba, neboť velká část webovských stránek se pravidelně mění.

Složky vektoru stránkového ocenění se v závěru výpočtu upraví pomocí funkce, která je tajná, a zaokrouhlují na celá čísla. Každá webovská stránka je tak ohodnocena číslem z množiny $\{0, 1, 2, \dots, 10\}$. Přitom samotná stránka `www.google.com` má ohodnocení 10. Stránka Matematického ústavu AV ČR a stránka JČMF mají shodné ohodnocení 6.

4. Závěr

Googlovský model oceňování webovských stránek je založen na jednoduchých počátečních předpokladech, které bohužel nevedou k dobře definovanému stránkovému ocenění. Proto bylo třeba přizpůsobit původní model tak, aby byla zaručena stochastičnost a kladnost výsledné googlovské matice, což pak umožnilo jednoznačně definovat vektor stránkového ocenění. Navíc rychlost konvergence mocninné metody je řádu $\mathcal{O}(\alpha)$, neboť největší vlastní číslo $\lambda_1 = 1$ je jednoduché. Druhé největší vlastní číslo v absolutní hodnotě je téměř jistě rovno teleportačnímu parametru $\lambda_2 = \alpha$.

Od roku 1960 začal postupně narůstat počet rozmanitých zařízení pro elektronické vyhledávání informací, která využívala zejména booleovské či pravděpodobnostní modely. Později se začaly používat i další modely, opírající se o vektorové prostory a fuzzy množiny [3, s. 5]. V roce 2000 bylo na internetu již 3500 různých vyhledávačů, z nichž každý měl poněkud odlišnou vyhledávací strategii. V silné konkurenci mezi nimi jasně zvítězil Google. V roce 2003 používal 15 000 počítačů a v současné době zhruba o řád více.

Úspěch internetového vyhledávače Googlu spočívá v tom, že dává nejvíce relevantní odpovědi na váš dotaz a je nesmírně rychlý. Pomáhá uživatelům i jako převodník jednotek a římských čísel, kalkulátor apod. Jeho efektivitu výstižně charakterizuje výrok známého amerického producenta Garryho Trudeaua: *Google is my rapid-response research assistant.*

Poděkování. Práce na tomto článku byla podpořena výzkumným záměrem AV0Z 101 90503 a projektem MŠMT č. 1P05ME749. Autoři děkují doc. RNDr. C. Matyskovi, DrSc., Mgr. V. Pravdovi, Ph. D., Mgr. Ing. J. Šolcovi a prof. RNDr. J. Wiedermanovi, DrSc., za podnětné připomínky.

L i t e r a t u r a

[1] Brin, S., Page, L.: *The anatomy of a large-scale hypertextual web searched engine.* Proc. of the 7th Internat. World Wide Web Conf., Brisbane 1998, 107–117.

[2] Golub, G. H., van Loan, C. F.: *Matrix Computations* (third edition). The John Hopkins University Press, Baltimore, London 1996.

[3] LANGVILLE, A. N., MEYER, C. D.: *Google's PageRank and Beyond: the science of search engine rankings.* Princeton University Press, Princeton, Oxford 2006.
[4] MAREK, I.: *Agregační variace na googleovskou matici.* Sborník semináře Matematika na vysokých školách (ed. HERRMANN, L.), Herbertov 2005, 114–120.
[5] MAREK, I., MAYER, P.: *Convergence theory of a class of aggregation/disaggregation iterative methods for computing stationary probability vectors of stochastic matrices.* Linear Algebra Appl. *363* (2006), 177–200.
[6] PAPPAS, T.: *More joy of mathematics.* Wide World Publ., Tetra 1991.
[7] PRAVDA, V., KŘÍŽEK, M.: *Citace: dobrý sluha, špatný pán.* PMFA *52* (2007), 28–36.
[8] `http://maps.google.com`
[9] `http://video.google.com`
[10] `http://www.google.com/help/features.html`
[11] `http://www.books.google.com`

Cambridgské setkání

Karel Žitný a Igor Zolotarev, Praha

V roce 1932 při pobytu na univerzitě v Cambridge Norbert Wiener (1894–1964) uspořádal sérii přednášek o Fourierově transformaci, které krátce nato shrnul do útlé, dodnes ceněné knížky „The Fourier Integral and Certain of its Applications“ [20]. Tento pozoruhodný text je jedním z trvalých výsledků, k nimž vedl jeho pobyt v Anglii; druhým, nikoliv však posledním, je článek [2] publikovaný G. H. Hardym (1877–1947). Tento matematik, starší než Wiener, byl autorem nebo spoluautorem mnoha skvělých statí a knih. Jeho sebrané spisy zaplňují sedm značně objemných svazků. Zaměříme-li nyní pozornost na jediný článek, není to bezdůvodné. Intelektuálně poctivý Hardy v něm podává svědectví o tom, jak významně jej ovlivnily diskuse s Wienerem a jak se podněty z neformálních rozhovorů mohou přetavit do korektně formulovaných tvrzení.

Oprávněně se můžeme domnívat, že vůdčím motivem, jenž vedl Wienera k tomu, že znovu přicestoval do Cambridge, byla touha po dalším prohloubení již existujících přátelských vazeb s tamější matematickou komunitou, zejména s Hardym. Pobyt byl úspěšný; nové renomé, které získal, mu rovněž umožnilo, aby navázal nové a oživil již fungující kontakty s matematiky z kontinentální Evropy. Avšak nejcennějším výsledkem se stala monografie [9]. Mimořádně talentovaný mladý cambridgský matematik Raymond E. A. C. Paley (1907–1933) získal Rockefellerovo stipendium a měl po Wienerově boku strávit akademický rok 1932–1933 na Massachusetts Institute of

---

RNDr. KAREL ŽITNÝ, CSc. (1936), Kamenická 5, 170 00 Praha 7.
Ing. IGOR ZOLOTAREV, CSc. (1954), José Martího 375/4, 162 00 Praha 6, e-mail: igor@it.cas.cz