Acta Informatica Pragensia, 2015, 4(3): 318–325
DOI: 10.18267/j.aip.78
Peer-reviewed paper

Několik myšlenek k tvorbě hesel

A Few Ideas for Creating Passwords

Petr Strossa[*], Radomír Palovský[*]

Abstrakt

Osmiznakových řetězců složených z malých a velkých písmen české abecedy a číslic existuje zhruba 6×10^{15}. Drtivou většinu z takových hesel si však nikdo není schopen zapamatovat, protože se ani vzdáleně neasociují s žádným „rozumným“ obsahem. V tomto textu docházíme k odhadu, že počet smysluplných českých vět o čtyřech až pěti slovech je určitě o několik desítkových řádů vyšší (a nezáleží přitom příliš ani na tom, zda použijeme nějaké velké písmeno), přičemž takto tvořená hesla se navíc dají snadno zapamatovat. Dále ukazujeme několik jednoduchých způsobů, jak rozšířit „prostor“ takto tvořených hesel až k hranicím kolem 10^{40} teoreticky možných řetězců, aniž by se pro uživatele příliš zkomplikovala zapamatovatelnost zvoleného hesla. Tím nabízíme metodu efektivní tvorby silných hesel.

Klíčová slova: Heslo, čeština, abeceda, gramatika, bezpečnost.

Abstract

There is about 6×10^{15} eight-character strings from Czech small and capital letters and numbers. The vast majority of such passwords is impossible to remember because of no association with any “reasonable” contents. In this paper we come to an estimate that the number of meaningful Czech sentences containing 4–5 words is certainly by several decimal orders higher (even without distinguishing small and capital letters), and passwords created in this way are easy to remember. Further we show some simple ways to extend the “space” of such passwords up to ca. 10^{40} theoretically possible strings without significantly complicating the possibility to remember the chosen password. A method for efficient generation of strong passwords is thus offered.

Keywords: Password, Czech language, Alphabet, Grammar, Security.

1 Úvod

Přestože autentizace uživatele pomocí hesla patří mezi nejstarší metody autentizace, je pořád jednou z hlavních metod, ne-li metodou pro běžné uživatele dominantní. Analýzou uživatelského chování, přístupu uživatelů k tvorbě hesel a doporučeními pro tvorbu bezpečnějších hesel se zabývá poměrně dost prací, z poslední doby např. Stanton et al. (2005), Komanduri et al. (2011). Nicméně v českém prostředí je prací poměrně málo, na vědecké

[*] Department of Information and Knowledge Engineering, Faculty of Informatics and Statistics, University of Economics, Prague, nám. W. Churchilla 4, 130 67 Praha 3, Czech Republic
✉ kizips@vse.cz, palovsky@vse.cz

úrovni téměř žádné, a navíc se zabývají spíše sociálními aspekty tvorby hesel (Páral, 2006) (Kadlecová, 2011). Naše práce se věnuje analýze složitosti vytvářených hesel, která vzhledem k jinému lingvistickému typu českého jazyka je tu významně jiná než v případě angličtiny.

V práci vycházíme jak z vlastních zkušeností s tvorbou hesel k různým účelům v různých prostředích, tak z dosud publikovaných prací o některých statistických vlastnostech češtiny (Čermák et al., 2004) (Hajič, 1996) (Hajič & Drozd, 1990) (Těšitelová et al., 1983) (Těšitelová et al., 1987). S využitím těchto zdrojů ukážeme, že *silná hesla* není nutné vytvářet v podobě *x*-znakových řetězců kombinujících velká a malá písmena s nejrůznějšími jinými symboly a nezapamatovatelných, protože nedávají žádný smysl, ale je možné vyjít z gramatických pravidel přirozeného jazyka. Ukážeme, že takto vytvořená hesla mají přinejmenším srovnatelnou sílu, a navíc je snadné si je zapamatovat.

2 Obecná pravidla tvorby hesel

Pravidla tvorby přístupových hesel k různým službám, ve kterých „o něco jde“, dnes nejčastěji vypadají nějak takto: heslo musí mít alespoň 8 znaků (nebo ještě lépe 12–15 znaků), v tom alespoň jedno malé písmeno, alespoň jedno velké písmeno a alespoň jednu číslici. Jinými slovy: minimalisticky utvořené heslo je osmiznakový (nebo až patnáctiznakový) řetězec z „abecedy“ malých písmen (tj. v češtině 42 znaků), velkých písmen (dalších 42 znaků) a číslic (10 znaků), tj. celkem 94 znaků. Takových hesel lze technicky vytvořit (a narušitel, který by chtěl metodou hrubé síly na heslo přijít, by jich musel zkusit vygenerovat) celkem $94^8 \approx 6 \times 10^{15}$ v osmiznakové variantě, resp. $94^{15} \approx 4 \times 10^{29}$ ve variantě patnáctiznakové.

V některých systémech ovšem může docházet k problému s rozpoznáním nastaveného typu klávesnice v okamžiku zadávání hesla. Pokud by se například v daném prostředí často střídalo používání anglické a české klávesnice, pak by asi nebylo příliš vhodné používat v heslech specificky česká písmena, ale ani písmena „y“ a „z“, která si vzájemně vyměňují polohu na klávesnici. V takovém případě by nám zůstalo jen 24 malých a 24 velkých písmen + 10 číslic, dohromady 58 znaků — a výslednou variabilitu možných hesel by vystihovala čísla $58^8 \approx 10^{14}$ v osmiznakové variantě, resp. $58^{15} \approx 3 \times 10^{26}$ ve variantě patnáctiznakové.

Ještě větší bezpečnost by samozřejmě mohla zajistit *ještě delší* hesla vytvořená podle obdobného schématu — jenže takové heslo (tvořené libovolnou kombinací velkých a malých písmen a číslic) je zpravidla už při méně než 8 znacích dost obtížně zapamatovatelné. A musíme-li si ho někam zapsat, pak jsme právě celou bezpečnost „poslali do háje“.

Považujeme za mnohem přirozenější nestarat se příliš o velká písmena ani o číslice, ale tvořit hesla ve formě *vět nebo frází přirozeného jazyka* o nějaké rozumné délce (nejde o nic jiného, než aby se člověk *příliš* nezdržoval jejich psaním) — jako např. „můj dědeček nebyl kosmonaut“…

3 Několik základních statistických charakteristik českého jazyka

Heslo výše uvedené jako příklad má konkrétně 27 znaků a nemáme problém si je zapamatovat. Všech možných řetězců o 27 znacích složených z písmen české abecedy (jen malých, tedy z abecedy 42 znaků) *a mezer* (bez jakýchkoli jiných interpunkčních znamének) je mimochodem $43^{27} \approx 10^{44}$. Vlastně ale příliš nezáleží ani na tom, zda mezi slovy používáme mezery. Řetězec „můjdědečeknebylkosmonaut“ se skládá z 24 malých písmen. Takových řetězců by technicky mohlo být vytvořeno $42^{24} \approx 10^{39}$. Musíme ovšem předpokládat, že

i potenciální narušitel naší bezpečnosti chápe, že když používáme tak dlouhá hesla (a nikam si je nepoznamenáváme, čili si je pamatujeme), pak nemůže jít o *libovolné* řetězce znaků, ale nejspíš půjde výhradně o řetězce dávající nějaký smysl, tj. o výrazy (fráze nebo věty) složené ze slov podle nějakého slovníku a gramatiky. Skutečně použitelných řetězců tedy bude *o něco* méně. Jsme schopni to aspoň trochu přesněji odhadnout?

Vyjděme z principu, že heslo má být větou nebo frází složenou alespoň ze čtyř slov. Je pravda, že věta ze čtyř slov v češtině může být třeba „a je i to“ — to je celkem 9 znaků včetně tří mezer — ale potenciální narušitel nemůže předem vědět, jak dlouhá slova jsme použili! Průměrně je třeba počítat s tím, že české slovo v textu má kolem 5,5 až 6 písmen (Těšitelová et al., 1987) čili věta nebo fráze o čtyřech slovech má kolem 25–27 znaků včetně mezer, resp. 22–24 znaků bez mezer.

Extrémně dlouhá slova mohou v češtině mít i hodně přes 30 písmen — ovšem taková slova, jako třeba „nejnezpravděpodobnostňovávatelnější“ (Těšitelová et al., 1987), asi nikdo ve svém hesle nepoužije, už proto, že je téměř nemožné napsat je na první pokus bez chyby, o možnosti napsat je správně „naslepo“ ani nemluvě… Za reálné však můžeme považovat použití *některých* slov dlouhých až kolem 10 písmen (přesně tolik má např. slovní tvar „kosmonauti“), celý výraz složený aspoň ze čtyř slov pak klidně může představovat i kolem 30–40 znaků, přičemž pro někoho může být snadno zapamatovatelný i výraz o dost delší. To jsme ovšem stále nezohlednili, že se ten výraz musí skládat ze slov, ne z libovolných písmen a mezer.

Konstrukce heslových frází místo heslových slov se přirozeně řeší i v angličtině. Významnou odlišností angličtiny od češtiny je téměř absentující tvarosloví a v důsledku toho obvykle gramatická smysluplnost náhodně vygenerované posloupnosti slov. Tohoto využívá např. systém *Diceware* (Reinhold, 1995) (Carnut & Hora, 2005), který vytváří heslové fráze náhodným výběrem slov. Korpus slov, který *Diceware* používá, tvoří přesně 7776 krátkých anglických slov. Číslo 7776 se rovná 6^5 a jednotlivá slova do fráze jsou vybírána hody klasickou šestihrannou kostkou (odtud název systému). Pět následných hodů vytvoří náhodné číslo, jemuž odpovídá jedno slovo. Doporučovaný způsob použití je vytvoření 5–8 slov podle typu použití hesla. Tím vznikne fráze, která fakticky reprezentuje 25 až 40 náhodných hodů šestihrannou kostkou a její složitost tedy je 6^{25} až $6^{40} \approx 10^{17}$ až 10^{27}. Nicméně čistě náhodné složení slov v češtině má problém v tom, že většinou netvoří smysluplnou frázi, zato díky tvarosloví máme mnohem více možností, čemuž se věnujeme dále.

4 Český slovník a gramatika (a co z toho plyne)

Kolik slov má čeština? To je „věčná otázka“, na kterou nikdy nebudeme znát přesnou odpověď — mimo jiné proto, že v každém lidském jazyce neustále (sice pomalu, ale jistě) vznikají nová slova a zanikají slova zastaralá. Nicméně lze odhadovat, že v současné češtině máme minimálně něco kolem 200 000 „obecně srozumitelných“ slov. Pokud k nim přidáme všechny vysloveně odborné termíny z nejrůznějších vědecko-technických oborů a všechna objektivně známá vlastní jména míst a osob, která lze v češtině použít, tento počet vzroste nejméně dvojnásobně, spíše však ještě mnohem více. (Hajič, 1996) (Těšitelová et al., 1983) (Těšitelová et al., 1987) (Záleží v podstatě jen na tom, jestli jsme připraveni použít jako součást našeho hesla např. jakékoli zeměpisné jméno z Ugandy. Zde je třeba si ještě uvědomit jednu důležitou věc. Nikdo z nás ve skutečnosti nedokáže používat 200 000 obecných slov ani 200 000 vlastních jmen. Individuální slovní zásoba člověka se pohybuje maximálně v oblasti desítek tisíc slov, u běžných lidí je spíše ještě o řád menší. Ovšem potenciální narušitel bezpečnosti nemůže vědět, která slova právě my známe a která ne. V tom je obrovská síla!)

Česká věta se ale neskládá „jen tak“ ze slov. Česká věta se skládá (podle jistých gramatických pravidel) ze *slovních tvarů* (podstatných jmen v určitém čísle a pádě, přídavných jmen v určitém stupni, rodě, čísle a pádě, sloves v určitém způsobu, času, osobě atd.). Různé slovní druhy mají různě široké arzenály možných tvarů. Tak například každé podstatné jméno existuje (*teoreticky*) ve dvou číslech a sedmi pádech, tj. ve 14 možných tvarech, a každé přídavné jméno může mít (spočítáme-li všechny možné stupně, rody, čísla a pády) dokonce 168 (!!!) různých tvarů. V praxi není různých tvarů slov tolik, jak by napovídal tento rozbor, jednak proto, že některá slova nelze použít ve všech tvarech (např. slovo „nůžky“ nemá tvary jednotného čísla, slovo „dvouhlavý“ logicky nelze stupňovat ap.), a dále proto, že prakticky u každého českého slova některé teoreticky rozlišované tvary znějí stejně (např. „nůžky“ je ve skutečnosti zároveň tvar prvního, čtvrtého i pátého pádu). V každém případě však nebudeme daleko od pravdy, budeme-li předpokládat, že v češtině máme k dispozici **více než milión (spíše asi několik miliónů) slovních tvarů** (Hajič & Drozd, 1990).

Kdyby tedy neexistovala žádná gramatická omezení na vzájemné kombinování různých slovních tvarů ve větě, existovalo by minimálně $(10^6)^4 = 10^{24}$ vět a frází o čtyřech slovech. Člověk by si samozřejmě mohl zvolit jako své heslo *libovolnou kombinaci čtyř slovních tvarů* — i když by to nebyla gramaticky správná kombinace (např. „čtvrtečnímu kolem nášlapná proti“) — jenže to bychom zase měli řetězec nedávající žádný smysl a z toho důvodu těžko zapamatovatelný. Zůstaňme tedy raději u představy *skutečné věty nebo alespoň fráze* (např. pojmenování něčeho) o 4 (nebo více) slovech. Případně připusťme i libovolné „výroky“, které přísně podle gramatiky nejsou správnými větami (něco v nich chybí ap.), ale které *intuitivně nějaký smysl dávají* (nebo aspoň naznačují), a proto se nám líbí a jsme schopni si je zapamatovat. Autoři tohoto textu mají například tento oblíbený „výrok“ popisovaného typu: „Jeden hnědý, druhý doleva.“

Ve skutečné smysluplné větě nelze slovní tvary kombinovat libovolně, to právě omezují pravidla *větné skladby* jazyka. Tato pravidla jsou ovšem poměrně velmi složitá a pro naši kombinatoriku mají velmi rozličné důsledky. Zpravidla platí, že n slov tvořících začátek věty do jisté míry omezuje volbu slova v pozici $(n + 1)$ — ale míra tohoto omezení dost záleží na tom, *jakých n* slov to je!

Jestliže například jako první slovo použijeme jakýkoli tvar podstatného jména, pak vzhledem k volnosti českého slovosledu není volba druhého slova prakticky vůbec omezena! A jen pro úplnost dodejme, že taková míra volnosti se zdaleka nemusí objevovat jen na začátku věty. Představme si například, že prvních $(n – 1)$ slov už tvoří jednoduchou větu a n-té slovo je spojka „a“. V takové situaci lze na pozici $(n + 1)$ pokračovat nanovo bez jakéhokoli omezení! (Délka „4 slova“ je poměrně malá, souvětí se spojkou „a“ se v této délce tvoří hodně těžko. Nicméně si připomeňme, že požadavek byl stanoven jako „alespoň 4 slova“, nikoli „přesně 4 slova“. Celkem nekomplikované souvětí může znít např.: „Teta přijela a zítra poletíme.“)

Na druhé straně, pokud jako první slovo použijeme přídavné jméno v 1. pádě jednotného čísla mužského rodu, dejme tomu „český“, pak se na první pohled jeví, že tím máme volbu následujícího slova poměrně dost omezenou: buď to bude další přídavné jméno, a to opět v 1. pádě jednotného čísla mužského rodu (např. „autorský“), anebo podstatné jméno mužského rodu, opět v 1. pádě jednotného čísla (např. „film“). Těžko přesněji odhadovat, ale ze všech možných tvarů všech českých slov by jich tato omezení mohla splňovat řádově snad desetina. Jenže ani tahle situace *na druhý pohled* tak přísná není — následujícím slovem smysluplné věty může ve skutečnosti být i leccos jiného, stačí vzít v úvahu možné úvodní fráze jako třeba „český žalostně nefunkční pseudotrh“ nebo „český ke všemu netečný politik“…

Zkusme proto — prozatím a čistě z nedostatku jiného, kvalifikovanějšího odhadu — předpokládat, že je-li v pozici číslo n ve větě použitelných $p(n)$ slovních tvarů, pak každá konkrétní volba *v průměru* omezí možnosti volby pro následující pozici na

$$p(n+1) = p(n) / 2$$

použitelných slovních tvarů. (Kdyby takové pravidlo skutečně platilo zcela obecně a přesně a čeština přitom dávala k dispozici právě jeden milión slovních tvarů, znamenalo by to, že 21. slovo ve větě je už s ohledem na předchozích 20 slov vždy jednoznačně dáno a 22 slov věta vůbec nemůže mít! V případě existence 4 miliónů slovních tvarů by se tato hranice posunula jen o 2 slova dál. Víme, že i delší smysluplné věty — tedy přesněji souvětí — odborníci dokážou napsat, takže náš předběžný odhad je snad stále „o něco přísnější než skutečnost“.)

Z výše uvedených předpokladů by vyplývalo, že **smysluplných vět a frází o přesně 4 slovech lze vytvořit přinejmenším** $1\,000\,000 \times 500\,000 \times 250\,000 \times 125\,000 \approx \mathbf{1{,}5 \times 10^{22}}$**.** Protože jsme na začátku stanovili, že se heslo má skládat *aspoň* ze 4 slov, nelze vyloučit, že si někdo vymyslí něco delšího. Celkový počet možností, se kterým je třeba počítat, by tedy měl být *přinejmenším* ještě asi 60 000 krát větší — to znamená, že by se měl pohybovat **kolem hodnoty** $\mathbf{10^{27}}$**.** I kdyby tedy potenciální narušitel měl k dispozici správnou generativní gramatiku spolu se správným slovníkem a pomocí těchto nástrojů zkoušel generovat možná hesla v podobě gramaticky správných frází a vět o délce v určitých mezích, vypadá tento přístup k tvorbě hesel „dost dobře“: jako uživatelé jsme schopni si taková hesla snadno zapamatovat, a jejich množství je přitom téměř srovnatelné s výše uvedenou maximalistickou hodnotu pro 15 libovolných znaků (přesněji by řád 10^{27} odpovídal 14 znakům).

Ale dokonce i kdyby se nakonec (na základě důkladného výzkumu kombinatoriky slovních tvarů podle české gramatiky) ukázalo, že náš odhad $p(n+1) = p(n) / 2$ byl příliš optimistický, je třeba uvést, že například pro variantu

$$p(n+1) = p(n) / 10$$

by při omezení na „alespoň 4 slova“ vycházela potřeba vyzkoušet minimálně kolem 10^{20} kombinací slovních tvarů (počítáme, že slov určitě může být i 5; $10^6 \times 10^5 \times 10^4 \times 10^3 \times 10^2 = 10^{20}$). I to je stále mnohem více než počet osmiznakových kombinací malých a velkých písmen a číslic (a to bez potřeby rozlišovat malá a velká písmena). Přitom absolutní platnost tohoto vzorce by teoreticky znamenala, že už pro sedmé až osmé slovo ve větě ve skutečnosti žádná možnost volby nezbývá a delší než osmislovné věty nejsou možné. Lze se tedy důvodně domnívat, že ani v průměru takto přísné omezení neplatí.

Nikomu přitom samozřejmě není třeba bránit v dalším posílení naznačeného schématu např. tím, že bude ve svém hesle dodržovat všechna pravidla interpunkce a psaní malých a velkých písmen (takže např. heslo může znít doslova: „Můj dědeček nebyl kosmonaut!“) — i když, popravdě řečeno, právě takovými detaily se ve skutečném přirozeném jazyce nic moc nezíská: velké písmeno má takhle krátká věta obvykle jen na začátku a nějaké interpunkční znaménko (z velmi omezené množiny) jen na konci.

Mnohem víc se dá naopak na síle hesla získat tím, že do něj cílevědomě zabudujeme nějakou chybu, kterou jsme schopni si zapamatovat — např. vynecháme jedno písmeno a použijeme formu „můj ddeček nebyl kosmonaut“. Případně můžeme, abychom si to všechno ještě bezpečněji pamatovali, aplikovat nějakou *systematickou chybu*, např. umělý „překlep“ typu záměny „ě“ ↔ „2“ (což se na české klávesnici skutečně často stává, takže se to „samo nabízí“) — a máme heslo ve tvaru „můj d2deček nebyl kosmonaut“. Nabízejících se systematických chyb podobného typu není sice přímo „nepřeberné“ množství, ale

potenciálnímu narušiteli bezpečnosti každopádně dost ztíží jeho snažení, že vůbec netuší, kolik a jakých podobných transformací jsme se rozhodli aplikovat.

Maximálně vhodné je taky kombinovat slova z různých jazyků, které trochu ovládáme — zhruba podle vzoru „byla jedna Mutter-matka“ ze známé dětské říkanky. V takovém případě se už vlastně počet „čtyřslovných“ konstrukcí, které by narušitel musel zkusmo generovat, blíží hodnotě x^4, kde x je úhrnný počet všech slovních tvarů všech v úvahu přicházejících jazyků, protože „slovní mix“ typu „byla jedna Mutter-matka“, byť *nám* dává určitý smysl, není generován gramatikou žádného jednotlivého jazyka!

Úplně nejlepší by asi bylo, kdybychom se podobně jako jistí američtí šifranti za 2. světové války naučili nějaký indiánský jazyk na prahu vyhynutí a hesla tolik potřebná pro náš dnešní *informatizovaný* život tvořili jako věty v tomto jazyce. Hackerovi, který by nevěděl, který jazyk jsme se naučili, by pak v podstatě nezbývalo než se smířit s tím, že pro heslo o délce např. 25 znaků (které my si snadno zapamatujeme a můžeme ho i různě průběžně obměňovat) existuje minimálně $27^{25} \approx 6 \times 10^{35}$ možných hodnot. (Tento výpočet vychází z předpokladu, že používáme 26 písmen mezinárodní abecedy a mezery mezi slovy.)

Protože však většina lidí zřejmě nemá tu možnost ovládnout jazyk, který by v jejich okolí neovládal nikdo jiný, můžeme jako „východisko z nouze“ nabídnout tuto poslední radu: do věty o 4–5 slovech lze zamontovat výraz, který známe, který nám dává smysl, ale který vlastně do žádného přirozeného jazyka přímo nepatří. Příkladně: „můj dědeček nebyl žádný e=mc2“. Takové výrazy je rovněž velmi těžké generovat gramatikou jakéhokoli konkrétního jazyka, o slovníku ani nemluvě, takže se opět — a tentokrát skoro „zadarmo“ — přibližujeme mezní situaci, kdy snažit se prolomit heslo určité délky znamená testovat všechny variace této délky ze všech možných znaků (přičemž zde konkrétně hovoříme o délce kolem 25–35 znaků vybraných ze souboru asi 60 znaků — pokud se nám stále nechce rozlišovat malá a velká písmena; to znamená, že existuje minimálně $60^{25} \approx 3 \times 10^{44}$ možných znakových řetězců).

5 Závěr

Opravdu silných hesel uživatel ve skutečnosti nepotřebuje mnoho. Značnou část použití hesel je možné ulehčit pomocí systémů klíčenek (password managers), které slouží jako uzamčené kontejnery pro používaná hesla. Jejich bezpečnost je někdy diskutabilní (Gasti & Rasmussen, 2012), ale i u těch bezpečných k otevření klíčenky obvykle heslo potřebujeme a jeho kvalita by měla odpovídat škodě, která by mohla nastat při neoprávněném prolomení. Další silné heslo je vhodné mít pro vlastní vstup do systému, když případná klíčenka ještě není k dispozici, nebo jako heslo, které je použito k odemčení klíčů použitých k šifrování disků a tím k přístupu k datům na těchto discích, pokud zájem na bezpečnosti a ochraně ukládaných dat významně přesahuje obtíže a zdržení pro stálé dešifrování.

Představili jsme zde několik úvah pro efektivní tvorbu silných hesel. Na základě jejich vyhodnocení, která výše uvádíme, soudíme, že by uživatelé nemuseli a neměli být nuceni k vymýšlení hesel ve formě náhodných sekvencí z poněkud uměle rozšířené množiny znaků, ale měla by být věnována větší pozornost (a to jak v praxi, tak v dalším teoretickém výzkumu) alternativní možnosti vytváření hesel v podobě frází přirozeného jazyka nebo podobajících se přirozenému jazyku. Podle našich dosavadních odhadů, které zde prezentujeme, mají taková hesla potenciál být silnější než dnes běžně používaná.

Seznam použitých zdrojů

Carnut, M. A., & Hora, E. C. (2005). Improving the Diceware Memorable Passphrase Generation System. In *Proceedings of the 7th International Symposium on System and Information Security*. São José dos Campos: CTA/ITA/IEC.

Čermák, F. et al. (2004). *Frekvenční slovník češtiny*. Praha: Nakladatelství Lidové noviny.

Gasti, P., & Rasmussen, K. B. (2012). On the Security of Password Manager Database Formats. In S. Foresti, M. Yung, & F. Martinelli (Eds.), *Computer Security – ESORICS 2012* (pp. 770–787). Berlin/Heidelberg: Springer. doi: 10.1007/978-3-642-33167-1_44

Hajič, J. (1996). Současnost a budoucnost inteligentní práce s textem. *ComputerWord*, (51–52), 25-40.

Hajič, J., & Drozd, J. (1990). Spelling-checking for highly inflective languages. In *Proceedings of the 13th conference on Computational Linguistics – Volume 3* (pp. 358–360). Helsinki: Association for Computational Linguistics.

Kadlecová, P. (2011). *Motivace uživatelů používat bezpečná hesla*. Bakalářská práce. Brno: Masarykova univerzita, Fakulta informatiky.

Komanduri, S. et al. (2011). Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2595–2604). New York: ACM. doi: 10.1145/1978942.1979321

Páral, K. (2006). *Hodnocení kvality hesel v počítačových systémech*. Bakalářská práce. Brno: Masarykova univerzita, Fakulta informatiky.

Reinhold, A. (1995). *Diceware Passphrase Home*. Retrieved from http://world.std.com/~reinhold/diceware.html

Stanton, J. M., Stam, K. R., Mastrangelo, P., & Jolton, J. (2005). Analysis of end user security behaviors. *Computers & Security*, 24(2), 124–133.

Těšitelová, M. et al. (1983). *Psaná a mluvená odborná čeština z kvantitativního hlediska (v rámci věcného stylu)*. Linguistica, IV. Praha: Československá akademie věd, Ústav pro jazyk český.

Těšitelová, M. et al. (1987). *O češtině v číslech*. Praha: Academia.