

**UNIVERZITA PARDUBICE  
ÚSTAV ELEKTROTECHNIKY A INFORMATIKY**

**EFEKTIVNOST NÁHLEDŮ STRÁNEK V INTERNETOVÉM  
KATALOGU**

**BAKALÁŘSKÁ PRÁCE**

**2007**

**Roman Honzák**

**UNIVERZITA PARDUBICE  
ÚSTAV ELEKTROTECHNIKY A INFORMATIKY**

**EFEKTIVNOST NÁHLEDŮ STRÁNEK V INTERNETOVÉM  
KATALOGU**

**BAKALÁŘSKÁ PRÁCE**

**AUTOR PRÁCE: Roman Honzák  
VEDOUCÍ PRÁCE: Mgr. Miroslav Suchý**

**2007**

**UNIVERSITY OF PARDUBICE  
INSTITUTE OF ELECTRICAL ENGINEERING  
AND INFORMATICS**

**THE EFFECTIVENESS OF THUMBNAIL PREVIEWS IN  
INTERNET CATALOGUES**

**BACHELOR WORK**

**AUTHOR: Roman Honzák  
SUPERVISOR: Mgr. Miroslav Suchý**

**2007**

**Vysokoškolský ústav:** Ústav elektrotechniky a informatiky  
**Katedra/Ústav:** Ústav elektrotechniky a informatiky  
**Akademický rok:** 2006/2007

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

**Pro:** Roman Honzák

**Studijní program:** Informační technologie

**Studijní obor:** Informační technologie

**Název tématu:** Efektivnost náhledů stránek v internetovém katalogu

### Zásady pro zpracování:

Vaším úkolem bude zjistit, zda internetové katalogy s náhledy stránek zvyšují efektivitu uživatele při hledání informací oproti katalogům bez náhledů stránek

### Seznam odborné literatury:

DZIADOSZ S., CHANDRASEKAR R., Do thumbnail previews help users make better relevance decision about web search result? In 25th annual international ACM SIGIR conference on Research and development in information retrieval  
Tempere, Finland, 2002, s. 365

WOODRUFF A., FAULRING A., ROSENHOLTZ R., MORRSIOM J., PIROLI P., Using thumbnails to search the Web in Proceeding of the SIGCHI conference on Human factors in computing systems,  
Seattle, Washington, United States, 2001, s. 198

**Rozsah:** Přibližně 30 stran

**Vedoucí práce:** Mgr. Miroslav Suchý

**Vedoucí katedry (ústavu):** prof. Ing. Pavel Bezoušek, CSc.

**Datum zadání práce:** 30. 11. 2006

**Termín odevzdání práce:** 12. 5. 2007

Prohlašuji:

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Univerzity Pardubice.

V Pardubicích dne 15. 08. 2007

Roman Honzák

## **PODĚKOVÁNÍ**

Chtěl bych poděkovat Mgr. Miroslavu Suchému za jeho pomoc a trpělivost při tvorbě této práce a poskytnutí webového prostoru na serveru Thumbnail.cz, bez kterého by tato práce byla jen těžko dokončena. Dále patří dík všem respondentům, kteří byli ochotní spolupracovat a našli si čas pro tento výzkum.

## **ABSTRAKT**

Tato práce se zabývá problematikou náhledů stránek v internetovém vyhledávači, konkrétně však jejich efektivností při hledání. Základním stavebním kamenem jsou naměřené časy, které uživatelé stráví hledáním v modifikovaném vyhledávači. Uživatel hledá odpovědi na otázky rozdělené do čtyř kategorií. Pro měření se využívají dva druhy odpovědí od vyhledávače. První nabízí pouze textovou nabídku vyhledaných odkazů, druhý k nim připojuje také jejich náhled. Předmětem výzkumu je dokázat, že náhledy v takové formě, v jaké se v současné době používají, mají pozitivní vliv na rychlost vyhledávání v internetovém katalogu a navázat tak na starší výzkumy.

## **ABSTRACT**

This work deals with problems of page view in Internet search engine, actually with their effectiveness in searching. Basic units are measured times that the users will spend while searching in forced search engine. User searches answers on questions divided into four categories. For measuring are used two sorts of answers from search engine. First offers only text suggestion of searched links, second incorporates also their preview. Subject of the research is to exemplified, that the previews which are currently used have positive influence over search speed in Internet search engine and follow up with previous researches.



# OBSAH

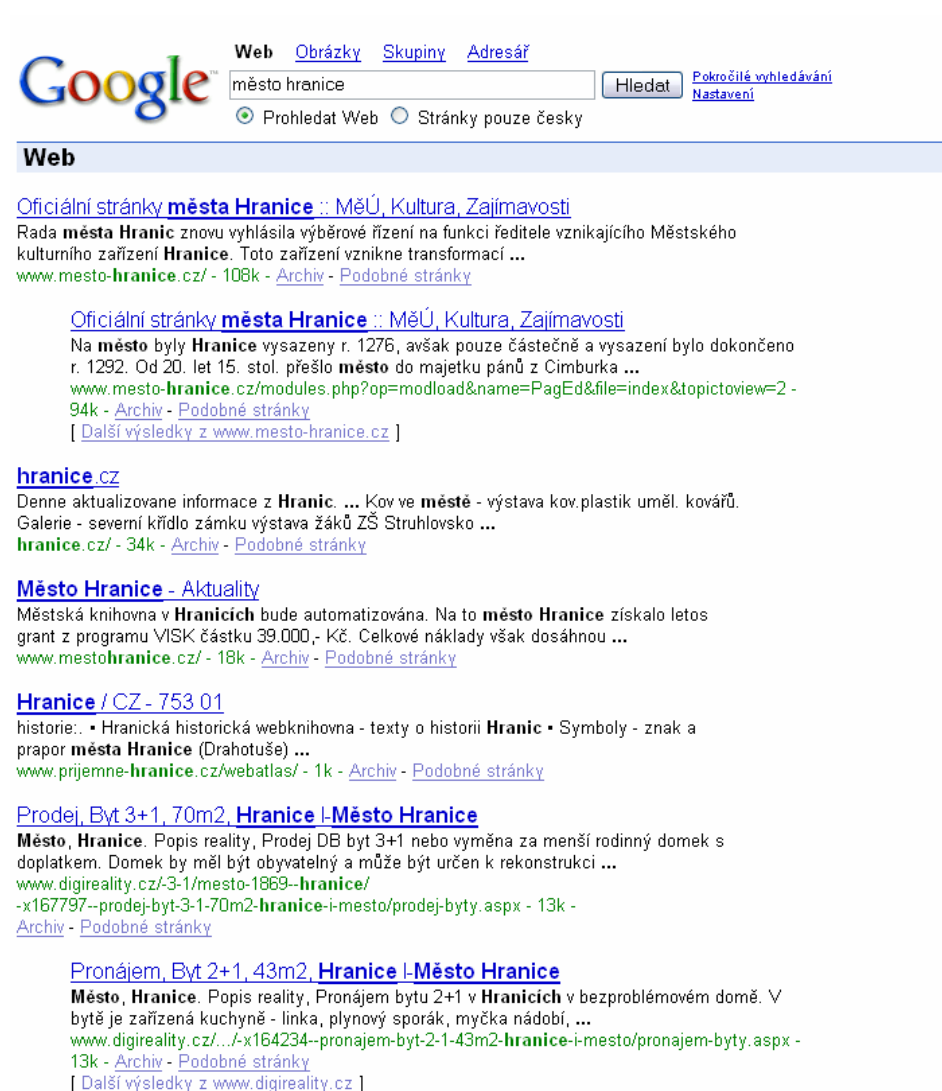
<b>1. ÚVOD</b>	<b>10</b>
<b>2. JINÉ VÝZKUMY</b>	<b>13</b>
2.1. DO THUMBNAIL PREVIEWS HELP USERS MAKE BETTER RELEVANCE	13
2.2. USING THUMBNAILS TO SEARCH THE WEB	13
2.3. V ČEM A PROČ NAVÁZAT?	14
<b>3. POPIS PROBLÉMU NÁHLEDŮ</b>	<b>15</b>
3.1. GENEROVÁNÍ A TVORBA NÁHLEDŮ	16
<b>4. POPIS APLIKAČNÍ ČÁSTI</b>	<b>17</b>
4.1. ULOŽIŠTĚ DAT	17
4.2. OKRUHY OTÁZEK	19
4.3. KLÍČOVÁ SLOVA PRO PLNĚNÍ ODKAZŮ	20
4.4. VELIKOST NÁHLEDU	22
4.5. METODIKA MĚŘENÍ	23
4.6. PROCEDURA DVANÁCTI OTÁZEK	24
4.7. VYHLEDÁVAČ	26
<b>5. CÍLOVÁ SKUPINA RESPONDENTŮ</b>	<b>27</b>
<b>6. VÝSLEDKY</b>	<b>28</b>
6.1. VÝSLEDKY PODLE SKUPIN RESPONDENTŮ	30
6.2. JEDNOTLIVÉ OKRUHY OTÁZEK	33
6.3. INDIVIDUÁLNÍ NÁZORY RESPONDENTŮ	37
<b>7. ZÁVĚR</b>	<b>38</b>

## SEZNAM OBRÁZKŮ A TABULEK

OBRÁZEK 1 - UKÁZKA VYHLEDÁVAČE BEZ NÁHLEDŮ <a href="http://WWW.GOOGLE.COM">WWW.GOOGLE.COM</a> .....	10
OBRÁZEK 2 - UKÁZKA VYHLEDÁVAČE S NÁHLEDY <a href="http://WWW.ALEXA.COM">WWW.ALEXA.COM</a> .....	12
OBRÁZEK 3 - NÁHLED INTERNETOVÝCH STRÁNEK <a href="http://WWW.NVIDIA.COM">WWW.NVIDIA.COM</a> .....	15
OBRÁZEK 4 – UML DIAGRAM TABULKY ODKAZY .....	18
OBRÁZEK 5 – UML DIAGRAM TABULKY VÝSLEDKY .....	18
OBRÁZEK 6 – DOTAZNÍK .....	24
OBRÁZEK 7 - PRŮBĚH MĚŘENÍ .....	25
OBRÁZEK 8 - MENU LISTOVÁNÍ VYHLEDÁVAČEM .....	26
OBRÁZEK 9 - VYHLEDÁVAČ S NÁHLEDY .....	26
OBRÁZEK 10 - VYHLEDÁVAČ BEZ NÁHLEDŮ .....	27
OBRÁZEK 11 - GRAF CELKOVÝCH VÝSLEDKŮ .....	28
OBRÁZEK 12 - GRAF VÝSLEDKŮ VYHLEDÁVÁNÍ TEXTOVÝM ODPOVĚDÍ .....	29
OBRÁZEK 13 - GRAF VÝSLEDKŮ VYHLEDÁVÁNÍ OBRAZOVÝCH ODPOVĚDÍ .....	29
OBRÁZEK 14 - GRAF CELKOVÝCH VÝSLEDKŮ ČLENĚNÝCH DLE ZNALOSTÍ .....	30
OBRÁZEK 15 - GRAF NAMĚŘENÝCH ČASŮ U ŽEN .....	31
OBRÁZEK 16 - GRAF NAMĚŘENÝCH ČASŮ U MUŽŮ .....	31
OBRÁZEK 17 - GRAF NAMĚŘENÝCH ČASŮ ŽEN, DĚLENÝCH DLE ZNALOSTÍ .....	32
OBRÁZEK 18 - GRAF NAMĚŘENÝCH ČASŮ U MUŽŮ, DĚLENÝCH DLE ZNALOSTÍ .....	32
OBRÁZEK 19 - SROVNÁNÍ ČASŮ KATEGORIE "OBRÁZEK, HLEDÁNÍ OBRAZOVÝCH ODPOVĚDÍ" .....	33
OBRÁZEK 20 - SROVNÁNÍ ČASŮ KATEGORIE "HOMEPAGE, HLEDÁNÍ DOMOVSKÝCH STRÁNEK" .....	33
OBRÁZEK 21 - SROVNÁNÍ ČASŮ KATEGORIE "E-COMMERCE, INTERNETOVÉ OBCHODY" .....	34
OBRÁZEK 22 - SROVNÁNÍ ČASŮ KATEGORIE "LÉKAŘSKÉ OTÁZKY" .....	35
OBRÁZEK 23 - SROVNÁNÍ ČASŮ KATEGORIE "LÉKAŘSKÉ OTÁZKY", POUZE TEXTOVÉ ODPOVĚDI .....	35
OBRÁZEK 24 - GRAF CELKOVÝCH VÝSLEDKŮ PRO JEDNOTLIVÉ KATEGORIE OTÁZEK .....	36
TABULKA 1 - VELIKOSTI NÁHLEDŮ NEJZNÁMĚJŠÍCH VYHLEDÁVAČŮ.....	22

# 1. ÚVOD

Snad každý, kdo použil někdy Internet se setkal s internetovým katalogem nebo vyhledávačem. Soubor dat, kterými tento stroj odpovídá na hledání, se ve velké míře jeví jako velmi obsáhlý a s přibývajícím množstvím internetových stránek stále nepřehlednější dokument.



The image shows a screenshot of a Google search results page. At the top, the Google logo is on the left, and navigation links for 'Web', 'Obrázky', 'Skupiny', and 'Adresář' are on the right. A search bar contains the text 'město hranice'. To the right of the search bar is a 'Hledat' button and a link for 'Pokročilé vyhledávání Nastavení'. Below the search bar, there are radio buttons for 'Prohledat Web' (selected) and 'Stránky pouze česky'. A blue horizontal bar with the word 'Web' is visible. The search results are listed below, each starting with a blue link to the official website of the town of Hranice, followed by a brief description and additional links like 'Archiv' and 'Podobné stránky'.

Web [Obrázky](#) [Skupiny](#) [Adresář](#)  
Google™ město hranice  [Pokročilé vyhledávání](#)  
[Nastavení](#)  
 Prohledat Web  Stránky pouze česky

**Web**

[Oficiální stránky města Hranice](#) :: [MěÚ](#), [Kultura](#), [Zajímavosti](#)  
Rada města Hranic znovu vyhlásila výběrové řízení na funkci ředitele vznikajícího Městského kulturního zařízení Hranice. Toto zařízení vznikne transformací ...  
[www.mesto-hranice.cz/](#) - 108k - [Archiv](#) - [Podobné stránky](#)

[Oficiální stránky města Hranice](#) :: [MěÚ](#), [Kultura](#), [Zajímavosti](#)  
Na město byly Hranice vysazeny r. 1276, avšak pouze částečně a vysazení bylo dokončeno r. 1292. Od 20. let 15. stol. přešlo město do majetku pánů z Cimburka ...  
[www.mesto-hranice.cz/modules.php?op=modload&name=PagEd&file=index&topictoview=2](#) - 94k - [Archiv](#) - [Podobné stránky](#)  
[ [Další výsledky z www.mesto-hranice.cz](#) ]

[hranice.cz](#)  
Denne aktualizovane informace z Hranic. ... Kov ve městě - výstava kov.plastik uměl. kovářů.  
Galerie - severní křídlo zámku výstava žáků ZŠ Struhlovsko ...  
[hranice.cz/](#) - 34k - [Archiv](#) - [Podobné stránky](#)

[Město Hranice - Aktuality](#)  
Městská knihovna v Hranicích bude automatizována. Na to město Hranice získalo letos grant z programu VISK částku 39.000,- Kč. Celkové náklady však dosáhnou ...  
[www.mestohranice.cz/](#) - 18k - [Archiv](#) - [Podobné stránky](#)

[Hranice / CZ - 753 01](#)  
historie: • Hranická historická webknihovna - texty o historii Hranic • Symboly - znak a prapor města Hranice (Drahotuše) ...  
[www.prijemne-hranice.cz/webatlas/](#) - 1k - [Archiv](#) - [Podobné stránky](#)

[Prodej, Byt 3+1, 70m2, Hranice I-Město Hranice](#)  
Město, Hranice. Popis reality, Prodej DB byt 3+1 nebo výměna za menší rodinný domek s doplatkem. Domek by měl být obyvatelný a může být určen k rekonstrukci ...  
[www.digireality.cz/-3-1/mesto-1869--hranice/-x167797--prodej-byt-3-1-70m2-hranice-i-mesto/prodej-byty.aspx](#) - 13k - [Archiv](#) - [Podobné stránky](#)

[Pronájem, Byt 2+1, 43m2, Hranice I-Město Hranice](#)  
Město, Hranice. Popis reality, Pronájem bytu 2+1 v Hranicích v bezproblémovém domě. V bytě je zařízená kuchyně - linka, plynový sporák, myčka nádobí, ...  
[www.digireality.cz/.../-x164234--pronajem-byt-2-1-43m2-hranice-i-mesto/pronajem-byty.aspx](#) - 13k - [Archiv](#) - [Podobné stránky](#)  
[ [Další výsledky z www.digireality.cz](#) ]

Obrázek 1 - Ukázka vyhledávače bez náhledů [www.google.com](http://www.google.com)

Prohlídnout celý tento soubor by pro uživatele bylo časově velmi náročné a vzhledem ke stereotypnosti dat (viz. Obrázek 1) i velmi nudnou záležitostí. Výrobci internetových katalogů a vyhledávačů na tuto skutečnost velmi dbají a zvyšují efektivitu hledání třemi způsoby.

- Vracení kvalitativně vyššího seznamu odkazů pomocí lepšího indexového pokrytí a lepších algoritmů pro seřazování.
- Vyhledávání ve specifikovaném jazyce
- Poskytnutí informace, která dovoluje uživateli rychlejší vyhodnocení výsledků vyhledávání.

Dnešní praxe se spíše ubírá směrem, poskytnout uživateli pouze krátký textový souhrn popřípadě logo, pokud je možné se do katalogu zaregistrovat jako firma. Textový souhrn obvykle obsahuje URL dokumentu, titulek, velikost, kategorii a několik vět, které buď shrnou dokument, nebo obsahují některé z vyhledávacích klíčových slov. Tyto souhrny mají výhodu malé velikosti, proto se velmi rychle stahují. Na druhou stranu, uživatel musí ohodnotit kvalitu informace čtením takového textového shrnutí. Čtení seznamů vyhledaných výsledků unaví, a empirická studia [1] dokazují, že průměrný uživatel vyhledávače je neochotný k tomu, aby přečetl víc než několik stránek takových výpisů. Jednoduché náhledy mají silné, ale také slabé stránky. Náhledy jsou typicky větší a proto se pomaleji stahují než textová shrnutí. Textový obsah v náhledech je těžce čitelný a není proto vhodný jako samostatná reprezentace výsledku vyhledávání [1]. Na rozdíl od toho, náhledy poskytují informaci o rozvržení, žánru, stylu stránky a jestli již někdy předtím uživatel tuto stránku viděl, nebo viděl jinou, téhle podobnou (viz. Obrázek 2).

The screenshot shows the Alexa search interface. At the top left is the Alexa logo with the tagline 'The Web Information Company'. A search bar on the right contains the text 'Fender Guitar' and a search button. Below the search bar, it indicates 'Results 1 - 10 of about 1,652,000 for Fender Guitar'. The main content area displays a list of search results, each with a small thumbnail image on the left and text on the right. The results include:

- Fender Electric Guitars**: Models in stock from the classics, American Deluxe, and Custom Shop. WildWestGuitars.com
- Amazon.com: "Fender Guitar": Books**: Amazon.com seeks to be Earth's most customer-centric company, where customers can find and discover anything they might. www.amazon.com/s?ie=UTF8&keywords=%22Fender%20Guit... - Cached. Rank: 29 | Site Info
- Amazon.com: Fender Guitar Head Official Necklace: Apparel**: Amazon.com seeks to be Earth's most customer-centric company, where customers can find and discover anything they might. www.amazon.com/Fender-Guitar-Head-Official-Necklac... - Cached. Rank: 29 | Site Info
- NME.COM - Win - Win a signed Fender guitar**: The online presence of the UK-based New Musical Express. www.nme.com/win240 - Cached. Rank: 5,106 | Site Info
- Genuine fender guitar parts, pickups, pickguards, clothing, t...**: Fender collectibles from 60th Anniversary items to Ashtrays. Our range... Uk online store specializing in Fender parts and accessories for the Fender Strat & ... On top of this we probably have the largest range of Fender T-Shirts ... www.jarsn-f.co.uk - Cached. Rank: 2,215,111 | Site Info
- Fender Guitar Strings from zZounds.com**: Fender Guitar Strings from zZounds.com ... www.zzounds.com/cat-Fender-Guitar-Strings--3079 - Cached. Rank: 9,215 | Site Info
- Fender Guitar**: General specifications, serial numbers, Fender vintage guitar cases... Top Web Resources for Fender Guitar... Vintage Fender Guitars - Overview of collecting vintage Fender guitars, basses and amps. www.axetopia.com/topics/fender\_guitar.html - Cached. Rank: 447,092 | Site Info

Obrázek 2 - Ukázka vyhledávače s náhledy [www.alex.com](http://www.alex.com)

Se zřetelem ke skutečnosti, že lidský vizuální systém může obraz zpracovat rychleji než text, může taková grafická informace mnoho úkolů ohromně zrychlit. Lidé můžou zpracovat podstatu obrazu ve 110ms [5]. V tomto čase člověk přečte průměrně méně než 1 slovo, nebo přeletí 2 slova. Průměrný čtenář angličtiny čte 4,2 slova za sekundu. Nepřekládá-li, může přelétnout zhruba 17 slov za sekundu. Je fakt, že hledání obrazu specifického objektu mezi dalšími obrázky je rychlejší než hledání jména tohoto objektu mezi dalšími slovy. Hypotéza této práce je tedy dokázat, že náhledy webových stránek pomohou uživatelům vyhledávačů k lepší orientaci ve výsledcích vyhledávání a tedy i rychlejšímu nalezení požadované odpovědi.

## **2. JINÉ VÝZKUMY**

### **2.1. DO THUMBNAIL PREVIEWS HELP USERS MAKE BETTER RELEVANCE**

Tento výzkum [2] pochází z dílny vývojářů Microsoftu. V tomto průzkumu se zkoumaly tři typy stránek: klasická odpověď od vyhledávače (titulek a krátký popis), pouze náhled stránky a klasický vzhled doplněný o náhled stránky. Autoři zjistili, že pokud jsou ve výsledcích pouze obrázky, tak uživatel potřebuje na posouzení menší mentální čas oproti pouze textové verzi (84,8 sec vs. 92,9 sec), ale pokud se k textové verzi přidají náhledy tak čas na posouzení o moc nevzroste (95,9 sec). Dále zjistili že kombinace náhled+text výrazně sníží počet nesprávných odmítnutí (uživatel si myslel, že stránka pro něj nebude přínosem, ale ve skutečnosti byla) a zvýší se počet správného výběru (uživatel si vybere stránku, která je pro něj přínosem. Poněkud překvapivé je zjištění že přítomnost náhledů zvýší počet uživatelem vybraných stránek, které chce dále prozkoumat (a na které klikne) o plných 15 %.

### **2.2. USING THUMBNAILS TO SEARCH THE WEB**

Tento výzkum [1] rozvojového centra firmy Xerox a Carnegie Mellon University se zabývá taktéž, jak je patrné z nadpisu, výzkumem vlivu náhledů při vyhledávání na internetu. Pro svůj výzkum použili tři druhy měření:

- Prostý text
- Obyčejní náhled
- Modifikovaný náhled

Modifikovaný náhled je velkou zajímavostí a skrývá velké množství modifikací, které ve výzkumu byly použity. Před vytvořením náhledu, preprocesor mění ve stránce barvu nebo velikost jistých elementů.

Při obdržení HTML dokumentu spojeného s danou URL, preprocesor odhadne zobrazení HTML elementů. Implementace softwaru je kompletně v jazyce Java. Uživatel poté specifikuje požadovaná upravení pomocí asociativního seznamu výrazových/stylových párů. Např. každé klíčové slovo bude zvětšeno na velikost písma 16pt nebo každý nadpis h1 bude mít žluté pozadí. Nevýhodou je, že vygenerování takového náhledu je výpočetně velmi náročné a zabere až několik sekund.

Hlavním přínosem tohoto výzkumu je zjištění, že vyhodnocení některých typů dotazů je velmi rychlé pomocí náhledů. Např. na prohledání výsledků dotazu typu „Najdi obrázek žirafy v divočině“ potřebovali uživatelé čtvrtinový (!) čas než u textové verze. Na dotazy typu „Najdi obchod, kde mohu koupit DVD přehrávač a zjisti jeho cenu“ potřebovali uživatelé pomocí náhledů pouze 3/4 času oproti textové verzi. Na druhou stranu existují skupiny dotazů, kde výsledky s náhledy podávají horší časy. Např. dotaz typu „Zjisti vedlejší efekty léku XY“ podávají náhledy o cca 10% horší výsledky a u dotazů typu „Najdi domovsku stránku Jana Nováka“ jsou horší o 54%.

### **2.3. V ČEM A PROČ NAVÁZAT?**

Tato práce se snaží volně navazovat na oba výzkumy a zkombinovat jejich metodiky v jednu. Oba dva průzkumy jsou již přes pět let staré. Výzkum Microsoftu (viz. 2.1) zkoumal stejné druhy náhledů jako tato práce, ale nedělil je do kategorií podle druhu otázek ani respondentů. V případě výzkumu Xeroxu (viz. 2.2) naopak rozdělení dle skupin bylo, ale náhledy, které byly použity, se dnes již nevyskytují a dokonce modifikovaný náhled se nikdy veřejně nepoužíval.

### 3. POPIS PROBLÉMU NÁHLEDŮ

Náhled lze chápat jako miniaturní zmenšeninu internetových stránek v podobě malého obrázku. Existují různé typy náhledů, např. náhled celé stránky, levého nebo pravého horního rohu, zmenšenina, nebo tzv. screenshot. Tyto typy se běžně vyskytují na internetu, ovšem některé další se objevují pouze ve výzkumných nebo testovacích pracích. Pro jejich složitost, respektive náročné generování se od běžného používání prozatím upustilo. Firma Xerox pro svůj výzkum [1] vyvíjela náhled, který zvýrazňoval vyhledaná klíčová slova na stránce a přitom vzhled hlavní rozložení stránky zůstalo neporušeno (viz. Kapitola 2.2). Tento náhled dnes v běžném vyhledávači na Internetu nenajdeme.

V hojném počtu se na místě náhledů objevují loga spojená s webovou prezentací. Bohužel, neexistuje žádný, veřejně dostupný, výzkum, který by potvrdil jejich přínos pro vyhledávání. Proto o jejich právoplatném umístění nezbývá, než spekulovat.

Mimo výše popsané nesporné výhody má náhled také velkou nevýhodu a tou je velikost. Náhled níže (viz. Obrázek 3) má velikost 3107 bytů, což je relativně 3107 znaků čistého textu, pokud je v kódování 8 bitů na znak. To vystačí zhruba na 12 textových odkazů včetně HTML formátování. Uživatelé s pomalejším internetem, dnes považováno 256kbit/s download a méně, si na načtení jedné stránky výsledků s náhledy počkají nejméně 3 sekundy (počítáno 30 textových shrnutí s náhledy a 10kB informací navíc).



Obrázek 3 - Náhled internetových stránek [www.nvidia.com](http://www.nvidia.com)



### 3.1. GENEROVÁNÍ A TVORBA NÁHLEDŮ

Pro generování a uchovávání náhledů je využito služeb serveru thumbnail.cz. Systém generování funguje následovně. Používá se k tomu jádro Konqueroru, linuxového internetového prohlížeče, konkrétně modul KHTML a stránka se nechá vykreslit. Vykreslování probíhá do paměti, takže je celý proces velmi rychlý. Pouze pokud stránka obsahuje flash (na to je nutno projít DOM strukturu stránky a hledat příslušné tagy), tak je nutno nechat stránku vykreslit do virtuálního framebufferu, protože flashplugin nepoužívá QT a vykreslený flash se vyskytuje v jiné oblasti paměti. Pokud stránka obsahuje nějaké aktivní prvky (např. javascriptové dialogové okno), pak je to detekováno a pomocí rozšíření Xserveru je učiněn pokus toto okno automaticky „odkliknout“. Stránka je po vyrenderování odeslána na server, kde je zmenšena na požadované rozměry a uložena do databáze. Poté se náhled vkládá do HTML jako img tag a jako zdroj se uvede adresa kterou vygeneruje server Thumbnail.cz. Pro samotné generování, upozornění serveru Thumbnail.cz aby vygeneroval náhled, byl vytvořen skript využívající funkce PHP `file_get_contents()`, která pokusem o načtení URL požadovaného náhledu vyžádá jeho vygenerování a uloží do databáze. Systém je schopen vygenerovat přibližně šedesát tisíc náhledů denně, tudíž při větším počtu náhledů je možné, že se budou náhledy zobrazovat až za nějaký čas.

## 4. POPIS APLIKAČNÍ ČÁSTI

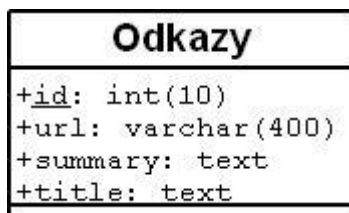
Celý proces měření využívá internetové aplikace vytvořené v jazyce PHP propojené s databázovým serverem MySQL ve formě meta vyhledávače. Výzkum probíhá v určitém časovém horizontu, proto je vyhledávač vytvořený tak, aby poskytoval všem respondentům výsledky ze stejné množiny odkazů. To je docíleno naplněním databáze určitým počtem odpovídajících odkazů, a výsledné odpovědi od vyhledávače jsou pouze v rozsahu odkazů v databázi. Celá aplikace spolupracuje s vyhledávačem [www.yahoo.com](http://www.yahoo.com) a tedy všechny odkazy pochází z tohoto zdroje. Pro analýzu výsledků Yahoo! je použito jejich datové služby API [4]. Tato služba umožňuje přijmout výsledky vyhledávání Yahoo! jako XML dokument a pomocí skriptu jej roztrždit a rozložit na jednotlivé elementy (URL, titulek, popis). Vše je možné provádět externě, mimo server Yahoo!.

Speciálně pro tuto práci byl vytvořen virtuální počítač, který funguje na serveru Thumbnail.cz s veřejnou IP adresou a DNS záznamem. Všechny operace spojené s prací se dějí pouze v rozsahu serveru Thumbnail.cz. Virtuální počítač je samostatný a slouží pouze tomuto projektu.

### 4.1. ÚLOŽIŠTĚ DAT

Pro účel uchovávání dat spojené s touto prací byla vytvořena výše zmíněná databáze. V ní jsou vytvořené celkem dvě tabulky, jedna pro uchovávání odkazů (viz. Obrázek 4), kde pro zamezení duplicit je na sloupci URL nastavený unique klíč, a druhá pro shromažďování konečných výsledků (viz. Obrázek 5). Pro bezproblémové zobrazování českých znaků byla zvolena v databázi znaková sada UTF-8 Unicode a pro porovnávání sada utf8\_czech\_ci, která nerozlišuje malá a velká písmena.

V tabulce Odkazy není záměrně sloupec náhled, protože náhledy jsou uloženy přímo na serveru Thumbnail.cz. Pro přístup k nim se vložením požadované URL stránky do připravené textové šablony vytvoří img tag, jehož zdrojem je odkaz na daný náhled uložený na serveru.



Obrázek 4 – UML diagram tabulky Odkazy



Obrázek 5 – UML diagram tabulky Výsledky

## 4.2. OKRUHY OTÁZEK

Podle systematické analýzy aktivit v Internetu, které významně ovlivňují rozhodování a činy lidí [3], jsou nejčastěji vyhledávanými okruhy tyto:

- Homepage, hledání domovských stránek
- E-commerce, internetové obchody
- Lékařské otázky

Tato práce bude využívat tohoto výzkumu a otázky byly vybrány z těchto tří okruhů. Pro očekávané lepší výsledky bude přidán pro porovnání jako čtvrtý okruh „Najdi fotku“, který bude mít svého zástupce i v okruhu „Lékařské otázky“. Otázky byly vybírány tak, aby respondentům nebyla ihned po položení otázky známa jejich přesná URL adresa nebo dokonce odpověď a aby obtížnost hledání pro každou entitu byla rovnoměrná rozložená. V okruhu „E-commerce“, internetové obchody“ byla zvolena nejčastější oblast nakupování v internetových obchodech a tou je elektronika. V okruhu „Najdi fotku“ byly otázky vybrány tak aby alespoň jedna byla zřetelně těžší než ostatní.

- Najdi fotku
  - Václava Havla
  - dvou Gripenů
  - havárie Formule 1
- Najdi homepage
  - Miroslava Suchého
  - hokejového klubu Vsetín
  - ski areálu Bílá

- Najdi cenu v internetovém obchodě
  - GeForce 8800 GTX
  - fotoaparátu Nikon D40
  - Blue-Ray disku
- Najdi
  - nežádoucí účinek Algifenu
  - nežádoucí účinek Neurolu
  - obrázek kuřákových plic

### 4.3. KLÍČOVÁ SLOVA PRO PLNĚNÍ ODKAZŮ

Skript, starající se o naplnění databáze potřebnými odkazy užívá pro jednu otázku více klíčových slov a jejich spojení. Nevyzpytatelnost respondentů v jejich formulaci pro vyhledávání činí tento úkol velmi obtížným. Slova musí být volena tak, aby zahrnula co největší počet odkazů pro danou otázku a respondenti měli z čeho vybírat. Preferovaným jazykem je samozřejmě čeština, protože lze s velkou pravděpodobností očekávat, že velká většina respondentů bude tento jazyk k vyhledávání používat. Ze stejného důvodu byla zahrnuta i diakritika, s tím, že každý uživatel bude před začátkem vyhledávání seznámen s charakteristikou vyhledávače a bude mu doporučeno používat diakritiku. Česká klíčová slova jsou skriptu zadávána v různých pádech a kombinacích, aby respondent nepřekvapil dotazem typu „fotka s dvěma gripeny“. Nelze ale tento problém vyřešit úplně, protože databáze, v tomto případě vyhledávače Yahoo!, jehož služby tento projekt využívá, není neomezená.

Jsou zahrnuty ale i slova, která databázi zaplní i cizojazyčnými odkazy. Zadávána v mezinárodních tvarech, kterým by měla při nejmenším angličtina rozumět.

Slova jsou také volena tak, aby se naplnily i takové odkazy, které nejsou pro nalezení vyhovující odpovědi příliš podstatná. Tím se simuluje reálný obsah Internetu, kde se ve většině případů nadá najít soubor výsledků, který by celý plně vyhovoval zadanému dotazu. Příklad klíčových slov pro dvě vybrané otázky:

- Otázka: Najdi fotku Václava Havla
  - Václav Havel
  - Havlův
  - Havlova
  - Havla
  - Havel
  - Václava Havla
  - fotka havla
- Otázka: Najdi cenu GeForce 8800 GTX v internetovém obchodě
  - GeForce 8800 GTX
  - GeForce 8800
  - GeForce 8800 prodej
  - GeForce 8800 cena
  - GeForce 8800 obchod
  - GeForce
  - nvidia

#### 4.4. VELIKOST NÁHLEDU

Velikost je velmi důležitým atributem náhledu. Pokud je náhled příliš malý, nejde dobře rozpoznat jeho obsah a respondent stráví mnoho času luštěním a později náhledy úplně ignoruje. Naopak pokud je příliš velký, váže na sebe veškerou pozornost a textové shrnutí je vedle něho ve většině případů úplně zbytečné. Optimální stav je takový, kdy si respondent z jednotlivých částí výsledku podle libosti vybere co mu v daný okamžik více vyhovuje či s kterou částí je více spokojen.

Velikost náhledů ve většině vyhledávačů je spojena s výškou textového souhrnu a téměř nikdy nepřesahuje jeho velikost. Při stejném pozadí odkazující stránky a vyhledávače rám jasně odděluje náhled od zbytku stránky, proto všechny zkoumané vyhledávače (viz. Tabulka 1) rám používají. Pro tento výzkum byla zvolena velikost náhledu 120x90 s černým orámováním pro jeho pěknou přehlednost a rychlé rozpoznání obsahu náhledu. Náhled je nestandardně vyšší než textový souhrn, to způsobuje větší šířka řádku souhrnu, kde je maximální počet pěti řádků pro jeden výsledek, z toho dva URL a nadpis.

Tabulka 1 - Velikosti náhledů nejznámějších vyhledávačů

URL vyhledávače	Velikost náhledu [v pixelech]
<a href="http://www.seznam.cz/">http://www.seznam.cz/</a>	100x75
<a href="http://www.atlas.cz/">http://www.atlas.cz/</a>	90x67
<a href="http://www.szm.sk/">http://www.szm.sk/</a>	100x75
<a href="http://www.golexa.com/">http://www.golexa.com/</a>	109x80
<a href="http://www.alex.com/">http://www.alex.com/</a>	112x82
<a href="http://www.exalead.com/">http://www.exalead.com/</a>	120x90

## 4.5. METODIKA MĚŘENÍ

Způsob získání času hledání byl předmětem důkladné analýzy, aby bylo dosaženo co nejlepších výsledků. Nejdůležitějším faktorem pro vyvážené a porovnatelné výsledky bylo určit, kde spustit čas a jak zhodnotit, zda respondent našel na otázku správný výsledek.

Zadat otázku, zapnout čas a nechat respondenta hledat se ukázalo jako nevhodné řešení, protože čas mezi zadáním otázky a spuštěním vyhledávání se u každého jednotlivce velmi liší. Váhání nad formulací klíčových slov pro vyhledávač a rychlost psaní na klávesnici jsou dva faktory velmi ovlivňující rozdíly časů ve fázi, která není pro tento výzkum natolik podstatná. Eliminací této části a tedy posunutí začátku měření až na okamžik, kdy respondent spustí vyhledávání, bylo dosaženo zvýšení kvality dosažených časů a výsledek zahrnuje pouze tu část, kdy je respondent opravdu v procesu vyhledávání. Opomenuta zůstává rychlost spojení. Všechny měření byly prováděny na počítačích s konektivitou nad 1 Mbit download a 512 kbit upload. Rozdíly časů, způsobené rychlejším připojením k Internetu, jsou zanedbatelné. V případě rychlosti zpracování dotazu serverem (analýza dat od Yahoo! API, porovnání s databází a zobrazení) je tomu obdobně, minimalizovat tento problém pomohlo zredukování maximálního počtu odpovědí od vyhledávače na pět set odkazů.

Za konec měřeného času se považuje okamžik, kdy si respondent myslí, že správnou odpověď našel i když nemusí být správná. Ověření zda-li je odpověď opravdu správná by značně skreslilo výsledný čas, protože by se do procesu zapojila druhá osoba. Čas, který je požadován má charakterizovat počin pouze testované osoby. Měří se tedy stopkami od momentu kdy je zmáčknuto tlačítko „hledej“ do té doby, kdy respondent našel uspokojující odpověď.



## 4.6. PROCEDURA DVANÁCTI OTÁZEK

Každý respondent prochází stejnou procedurou dvanácti náhodně po sobě jdoucích otázek ze čtyř kategorií s náhodným atributem přítomnosti náhledů. Pravděpodobnost dosažení vyhledávače s náhledy je 0,5.

Postup měření je následující. Jako první respondent vyplní formulář pravdivými osobními údaji (viz. Obrázek 6) a klikne na „Vygenerovat otázku“, tím se vygeneruje první otázka. Systém si hlídá nevyplněné políčka „Jméno“ a „Příjmení“. Pokud nejsou vyplněny zobrazí se červené upozornění pod formulářem a systém dále nepokračuje. Pokud se stane případ že respondent má problém se zveřejněním svých údajů vyplní místo jména a příjmení své iniciály. Pohlaví, věk a Znalosti jsou povinná pole, bez nichž by nešlo přiřadit výsledek odpovídající kategorii. Autor dohlíží na potřebné vyplnění formuláře. Pokud respondent nesouhlasí s vyplněním, výzkumu se nemůže zúčastnit.

Zde prosím vyplňte základní osobní údaje :

Jmeno *	<input type="text"/>
Prijmeni *	<input type="text"/>
Pohlaví	Muž <input type="button" value="v"/>
Věk	15 <input type="button" value="v"/>
Znalosti v oboru IT	Začátečník <input type="button" value="v"/>

\*) Tyto údaje nebudou nikde zveřejňovány nebo publikovány! Slouží pouze k identifikaci výsledků měření.

Obrázek 6 – Dotazník

Autor se zavazuje, že všechna sesbíraná data slouží pouze a jen pro účely této práce a nebudou dále nijak šířena či prezentována.

Stisknutím tlačítka „Otevři vyhledávač“ (viz. Obrázek 7) se v novém okně otevře vyhledávač v kterém je nutno provést hledání. V žádném jiném vyhledávači není pro tuto práci povoleno vyhledávat. Po změření času vyhledávání, odpovědná osoba, která celý proces také měří, vyplní položku „Výsledný čas“ a potvrdí. Vše se opakuje až po zodpovězení všech dvanácti otázek. Osobní data respondenta se přenáší celou procedurou a není třeba je znovu vyplňovat. Poté už systém nemá co nabídnout, neumožní další generování, a vypíše informaci i s poděkováním. Procedura je u konce. Pro jedno jméno je možné zodpovědět pouze dvanáct otázek. Při pokusu o další, systém odmítne.

Zde prosím vyplňte základní osobní údaje :

Jmeno *	<input type="text" value="Jan"/>
Prijmeni *	<input type="text" value="Novák"/>
Pohlaví	<input type="text" value="Muž"/>
Věk	<input type="text" value="22"/>
Znalosti v oboru IT	<input type="text" value="Profesionál"/>

\*) Tyto údaje nebudou nikde zveřejňovány nebo publikovány! Slouží pouze k identifikaci výsledků měření.

Vaše otázka č.3 zní : **Najdi fotku havárie Formule 1**

Vysledny cas :

**Obrázek 7 - Průběh měření**

## 4.7. VYHLEDÁVAČ

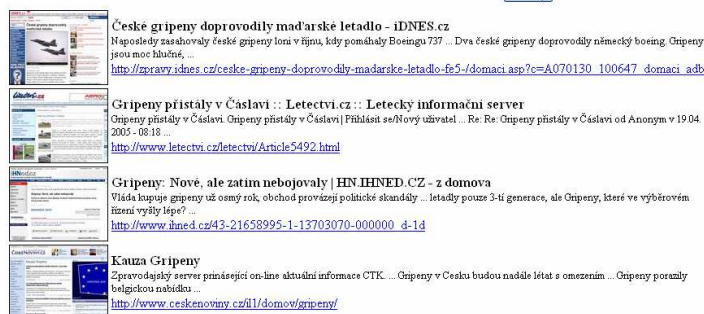
Nejdůležitější část práce, samotný vyhledávač, který spojuje všechny části této práce v jeden komplexní projekt. Vyhledávací mechanismus využívá technologie Yahoo! API [4], kde se hledaný řetězec posílá spolu s parametry, které určují atributy vrácených výsledků na Yahoo! a ve formě XML dokumentu se vrací zpět výsledky vyhledávání. Analyzují se pomocí PHP skriptu, jehož jádro je dílem vývojářů Yahoo!. Výsledné odkazy se porovnávají s databází uložených odkazů a ty, které jsou v ní obsaženy se zobrazí jako výsledek hledání. Tím je poskytnuto každému respondentovi shodné prostředí pro vyhledávání. Maximální počet zobrazených odkazů na stránce činí třicet. Respondent má možnost listovat mezi dalšími stranami výsledků (viz. Obrázek 8) až do maximálního počtu pěti set. Vyhledávač nabízí dva druhy výsledků, s náhledy (viz. Obrázek 9) a bez náhledů (viz. Obrázek 10).



Obrázek 8 - Menu listování vyhledávačem

### Bakalářská práce | Vyhledávač | Roman Honzák

Vaše otázka : Najdi obrázek dvou Gripenů

Obrázek 9 - Vyhledávač s náhledy

Vaše otázka : Najdi homepage ski areálu Bílá

SKI Vitkovice Bílá  
Provoz vždy So, Ne a svátky v 10,00 a 13,00hod. ... ? všechna práva vyhrazena SKI Bílá, Uvážené informace mají  
informativní charakter ...  
<http://www.skibila.cz/>

ski mezivodi  
areál je známý výbornými sněžovými podmínkami po celou sezónu ( prosinec - duben ) ... speciální SKI CROSS pro  
snowboard i lyže ...  
<http://www.skamezivodi.cz/>

Ski areál Bílá - Turistik.cz  
Přehledný průvodce po místech České republiky. Zajímavé informace o atrakcích ... Okres Frýdek-Místek, Ski  
areál Bílá, Staré Hamry, Běláky, ... Ski ...  
<http://www.turistik.cz/cs/pruvodce/ci/ski-areal-bila/>

LYŽARSKÝ SKI AREÁL BÍLÁ - MEZIVODÍ  
obce ostravice, obec-ostravice, ostravice, dovolení v beskydech, beskydy, beskydeky, beskydská, lysá hora,  
galerie ... areál Ski Mezivodí se nachází v ... Ski ...  
<http://www.ostravice.multimedia.cz/view.php?cisloclanku=2007010003>

Obrázek 10 - Vyhledávač bez náhledů

## 5. CÍLOVÁ SKUPINA RESPONDENTŮ

Lidé, kteří byli vybráni pro tento výzkum, mají všichni alespoň základní znalosti práce s počítačem, muži i ženy do věku padesáti let. Souhlasili s poskytnutím potřebných osobních údajů. Všichni byli schopni ovládat vyhledávač vytvořený pro tento výzkum a dokázali v něm vyhledávat samostatně, bez cizí pomoci.

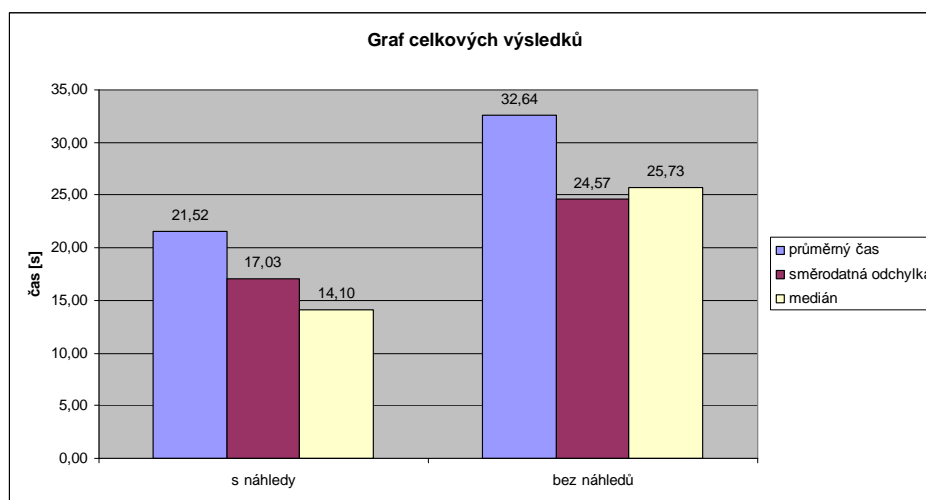
Respondenti byli také pomocí základních osobních údajů začleněni do skupin, aby bylo možno vypracovat charakteristické výsledky pro jednotlivé skupiny. Člení se dle následujících :

- Pohlaví
- Znalosti v oboru IT

Vybráno bylo čtrnáct lidí z toho čtyři ženy. Do kategorie začátečníci se zařadilo pět lidí z toho dvě ženy. Kategorii pokročilý obsadilo taktěž pět lidí z toho dvě ženy a jako profesionálové se uvedli čtyři muži.

## 6. VÝSLEDKY

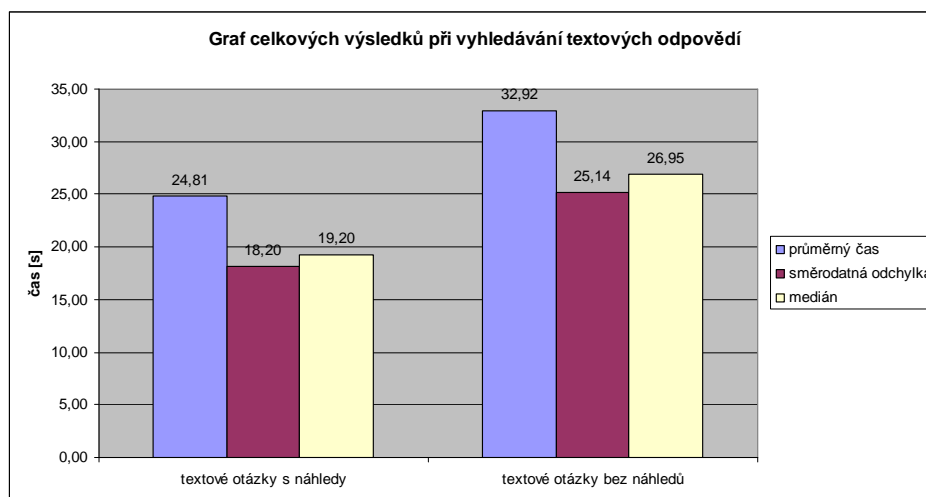
Tato kapitola bude obsahovat kompletní výsledky výzkumu, dělené do kategorií s komentářem. Výsledná čísla byly zaokrouhleny na dvě desetinná místa, pro všechny výsledky byl použit aritmetický průměr ochuzen o lokální extrém příslušné množiny čísel. Autor se zavazuje v pravost všech použitých zdrojových dat.



Obrázek 11 - Graf celkových výsledků

Celkové výsledky ukazují, že hypotéza práce se nezamítá a časy, dosažené s náhledy se ukazují v celkovém souhrnu jako lepší (viz. Obrázek 11). V tomto grafu jsou zahrnuty všechny výsledky, dosažené tímto výzkumem. Rozdíl dvanácti sekund vypovídá o nesporné důležitosti náhledů v internetových katalozích a jejich efektivní pomoci při hledání.

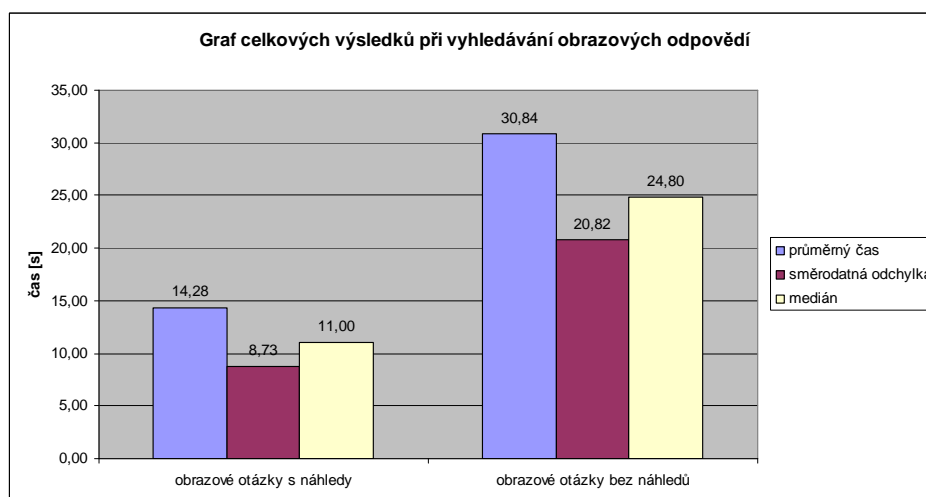
Tento graf se může zdát být neobjektivní z důvody přítomnosti otázek, kde respondent hledá jako odpověď obrázek nebo fotografii, jejichž lepší výsledky jsou velmi očekávány. Pro tento účel byl zpracován následující graf (viz. Obrázek 12), který je sestaven pouze z časů otázek, na které je textová odpověď.



**Obrázek 12 - Graf výsledků vyhledávání textovým odpovědí**

Ve výsledcích jsou zahrnuty pouze časy dosažené u otázek s textovou odpovědí. Oba časy narostly, bez náhledů mírně, s náhledy více. Jak je ale na první pohled patrné, převaha náhledů stále zůstává a rozdíl časů činí přes osm vteřin.

Pro srovnání je na následujícím grafu (viz. Obrázek 13) uveden rozdíl časů při hledání pouze obrazových odpovědí. Na tomto grafu je již vliv náhledů velmi patrný.



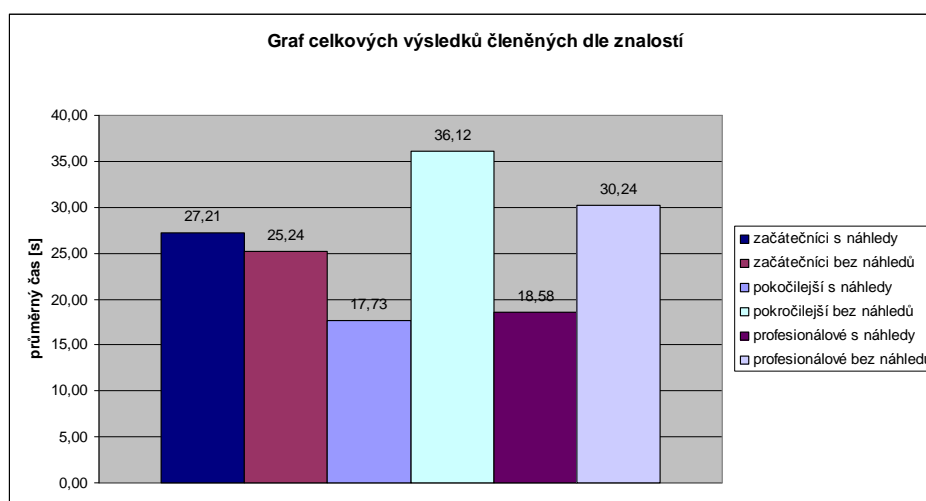
**Obrázek 13 - Graf výsledků vyhledávání obrazových odpovědí**

## 6.1. VÝSLEDKY PODLE SKUPIN RESPONDENTŮ

Jako poslední celkový souhrn je graf dělených respondentů dle jejich znalosti. Každý se v úvodním formuláři mohl, dle svého úsudku zařadit do jedné ze tří kategorií:

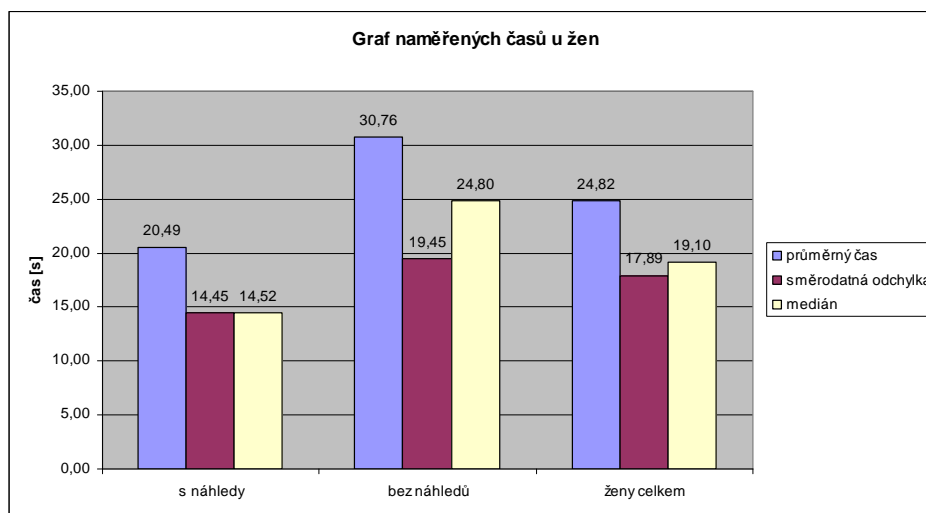
- Začátečník
- Mírně pokročilý
- Profesionál

Na základě tohoto rozdělení byl vypracován graf (viz. Obrázek 14), popisující časy dosažené v jednotlivých skupinách. Lze vyvodit závěr, že přítomnost náhledů opět zaručuje rychlejší nalezení dostačující odpovědi, výjimku tvoří kategorie začátečníci, jejichž zajímavý posuv oproti zkušenějším uživatelům, je dle mých vlastních pozorování respondentů, způsoben až příliš zmateným pohybem po stránce, kde si odkaz vybírali po přečtení souhrnu takřka náhodně. Ve dvou případech respondent rozčilen neúspěchem nalezení odpovídajícího odkazu zvolil metodu „Aááá tak třeba tenhle“. Z toho faktu lze vyvodit, že začátečníkům náhled k vyhledávání neposkytuje prakticky žádnou pomoc. Zajímavý je také fakt, že začátečníci dosahují bez náhledů lepších výsledků než obě skupiny.

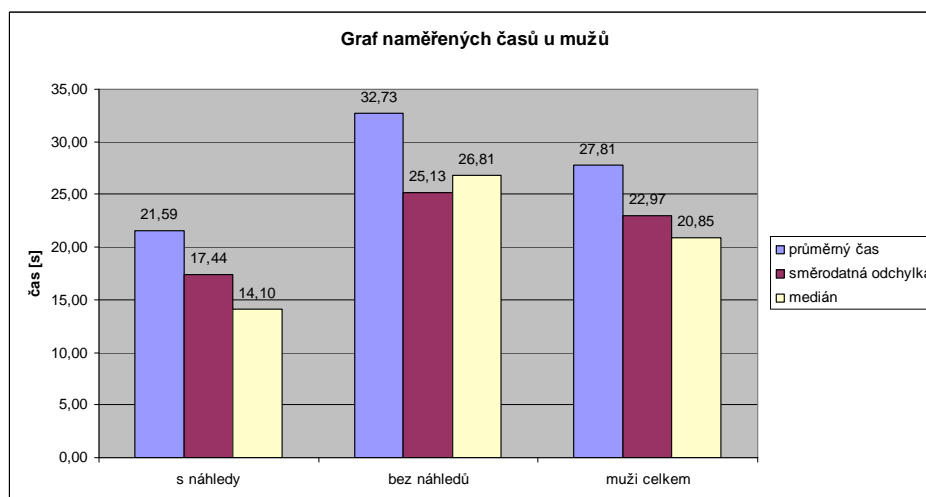


Obrázek 14 - Graf celkových výsledků členěných dle znalostí

Výsledky budou dále rozdělené jednotlivě do kategorií, do kterých byly respondenti zařazováni. Jako prvotní charakter pro skupinu respondentů bude považováno pohlaví. První graf je uveden celkový souhrn počínání ženského pohlaví (viz. Obrázek 15) a poté graf mužského pohlaví (viz. Obrázek 16). Ženy jsou na tom při hledání o málo lépe, orientují se lépe v textových souhrnech.



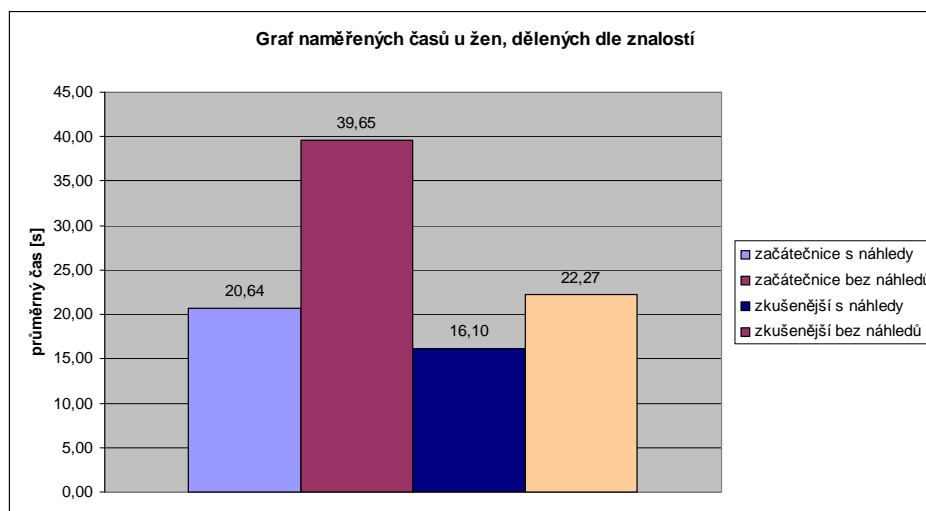
Obrázek 15 - Graf naměřených časů u žen



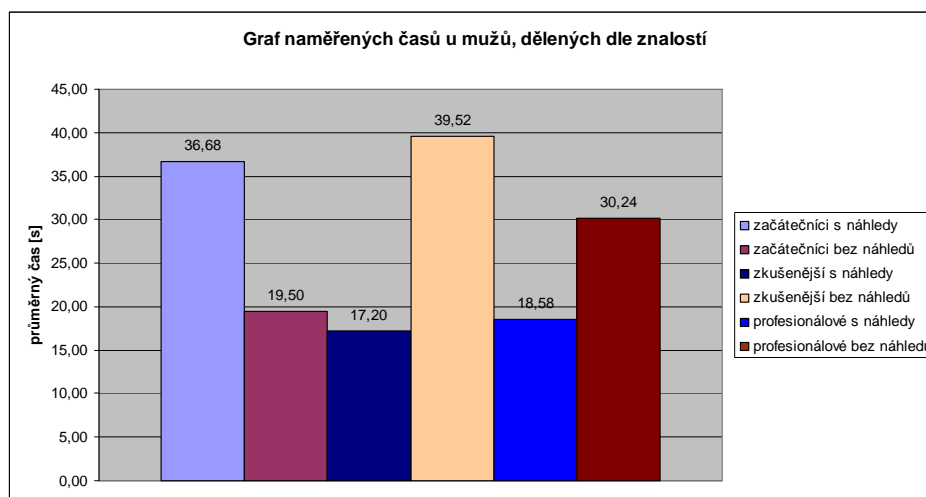
Obrázek 16 - Graf naměřených časů u mužů



Jako další parametr pro dělení je znalost práce s počítačem. U žen jsou dvě kategorie jasné (viz. Obrázek 17), kde ani jedna žena není v kategorii „profesionál“. U mužů se vyskytla zvláštnost v podobě lepších časů začátečníků bez náhledů. Tento fakt může být statisticky nepřesný, z důvodu malého počtu respondentů v kategorii „začátečníci“ (viz. Obrázek 18).



**Obrázek 17 - Graf naměřených časů žen, dělených dle znalostí**

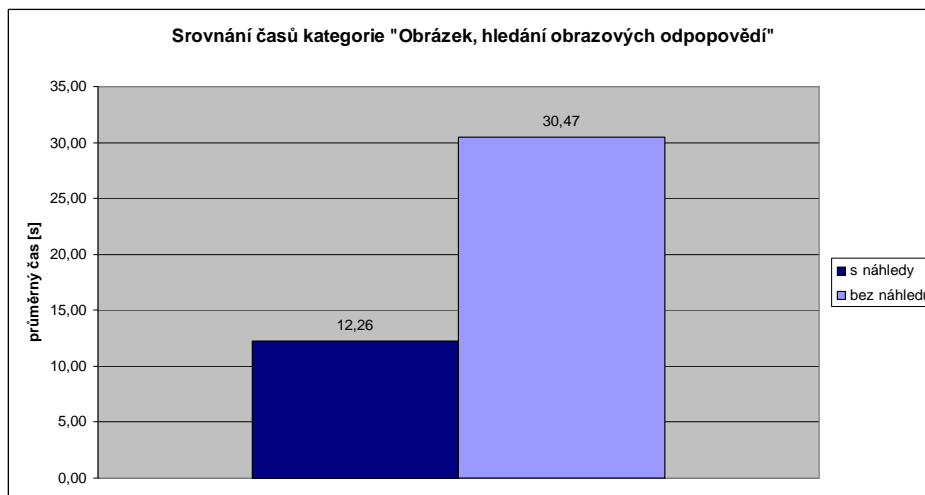


**Obrázek 18 - Graf naměřených časů u mužů, dělených dle znalostí**

V obou případech si lze povšimnout lepších časů u zkušenější kategorie respondentů. Náhledy prokazují lepší výsledky v obou kategoriích jak u mužů tak u žen.

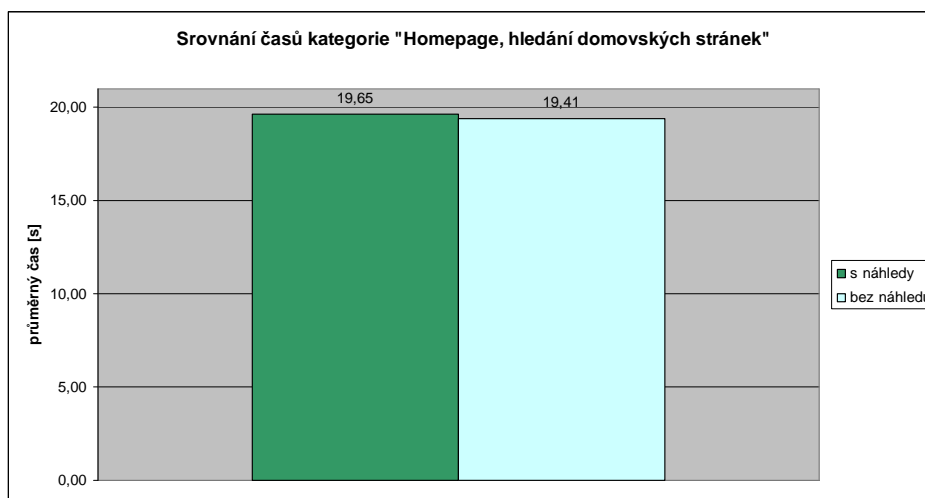
## 6.2. JEDNOTLIVÉ OKRUHY OTÁZEK

Výsledek první kategorie, který byl velmi očekáván a také nezklamal. V první kategorii hledali respondenti pouze obrazové typy odpovědí a graf (viz. Obrázek 19) opět dokazuje, že pro tento účel hledání je vizuální pomoc velkým přínosem.



Obrázek 19 - Srovnání časů kategorie "Obrázek, hledání obrazových odpovědí"

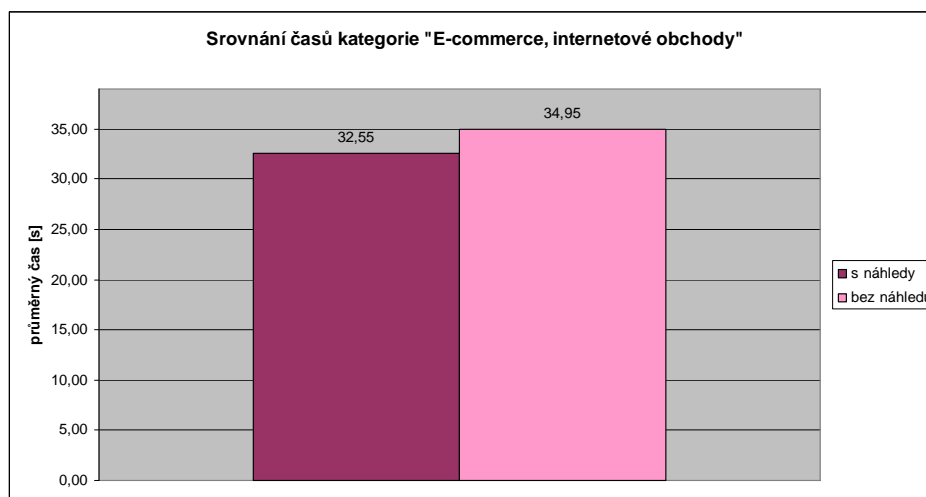
V druhé kategorii „Homepage, hledání domovských stránek“ už výsledek tak jednoznačný není, naopak výsledky jsou v porovnání téměř totožné (viz. Obrázek 20).



Obrázek 20 - Srovnání časů kategorie "Homepage, hledání domovských stránek"

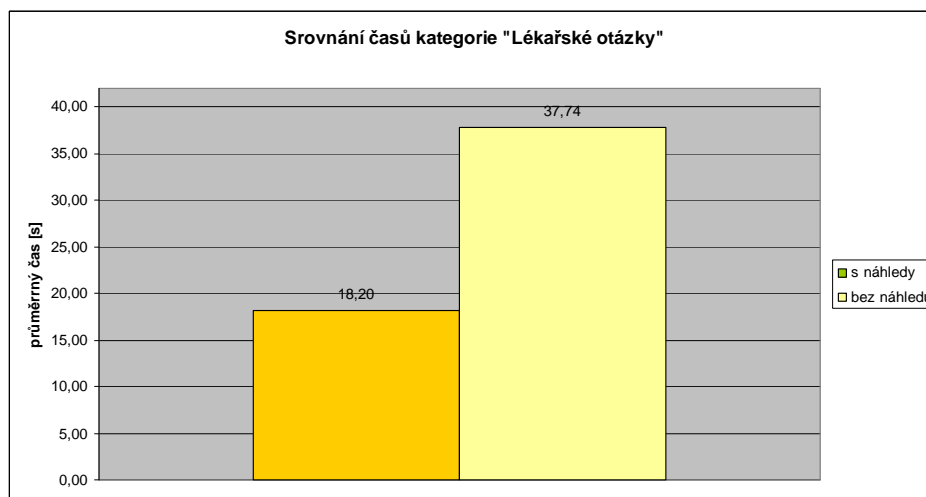
Tento fakt může být způsoben, že při hledání těchto typů otázek na internetu není náhled pro uživatele natolik důležitý. V případě domovské stránky bývá jednoznačně vypsáno textové shrnutí s výstižným popisem stránky a ještě výstižnějším nadpisem v podobě titulku stránky. Textové shrnutí je natolik výmluvné, že zkoumat náhled by bylo v okamžik, kdy uživateli takto podaná informace stačí k tomu aby si byl jist, že zrovna tato stránka je ta správná, je naprosto zbytečné. Samozřejmě princip funguje i opačně. Časy které respondenti dosahují v této kategorii při hledání bez náhledů, jsou relativně stejné jako dosahují u jiných kategorií s náhledy a z ostatních kategorií se průměrný výsledek tomuto ani nepřibližuje. Nejbliže je o jedenáct sekund horší výsledek z první kategorie. To vypovídá o značné jednoduchosti vyhledávání v této kategorii.

V kategorii třetí „E-commerce, internetové obchody“ je naopak jasně vidět obtížnost takového typu vyhledávání. Oba časy jsou jedny z nejhorších, čas s náhledy úplně nejhorší. I přes zjevnou obtížnost si náhled v této dosáhl lepšího výsledku, i když jenom nepatrného (viz. Obrázek 21).



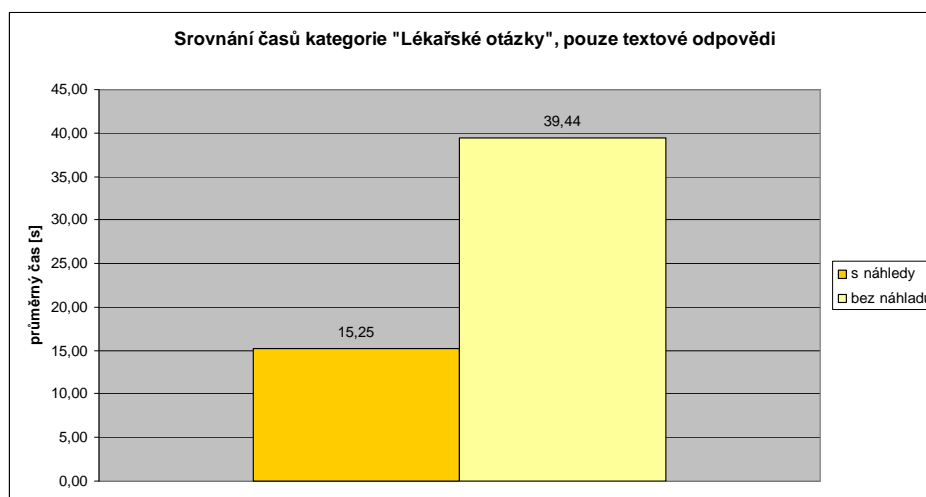
Obrázek 21 - Srovnání časů kategorie "E-commerce, internetové obchody"

Ve čtvrté a tedy poslední kategorii „Lékařské otázky“, je podle výsledků také očekáván dobrý výsledek, protože jedna ze tří otázek má obrazovou odpověď (viz. Obrázek 22). Čas s náhledem v této kategorii dosáhl rekordní hodnoty. V první kategorii, kde se všechny otázky zodpovídaly obrazově, dosahuje čas s náhledy o dvě a půl vteřiny horšího výsledku (viz. Obrázek 19).



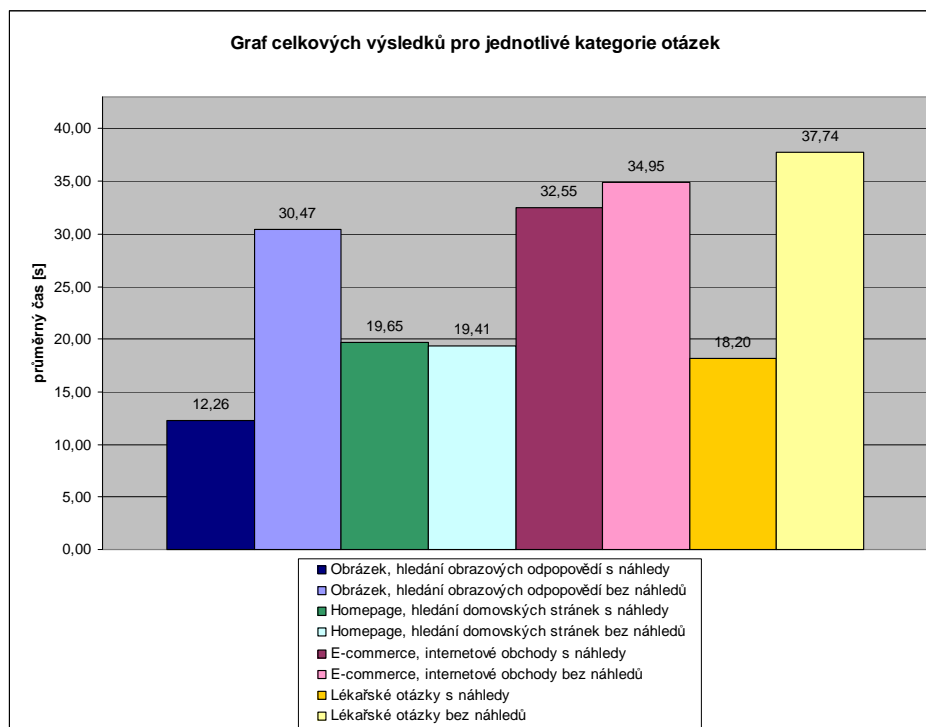
**Obrázek 22 - Srovnání časů kategorie "Lékařské otázky"**

Přítomnost jedné otázky s náhledem se může zdát jako faktor, velmi ovlivňující výsledek. Tuto spekulaci vyvrací graf (viz. Obrázek 23)



**Obrázek 23 - Srovnání časů kategorie "Lékařské otázky", pouze textové odpovědi**

Pro srovnání všech dosažených výsledků v jednotlivých kategoriích je uveden následující graf (viz. Obrázek 24), z něhož jsou opětovně jasně viditelné lepší výsledky vyhledávače s náhledy. Lze zde taky velmi snadno opticky porovnat složitost jednotlivých kategorií.



**Obrázek 24 - Graf celkových výsledků pro jednotlivé kategorie otázek**

### 6.3. INDIVIDUÁLNÍ NÁZORY RESPONDENTŮ

Při každém měření, které se konalo v osobním kontaktu s autorem práce byly zjišťovány názory, jak na detaily, tak i na celkový dojem z vyhledávání. Mnozí začátečníci si do této chvíle ani nepovšimli, že náhledy vůbec existují. Jejich funkce byla jednomu respondentovi zcela neznámá. Je to dáno tím, že u nás existuje pouze jeden jediný veřejně známý internetový katalog a vyhledávač s náhledy [www.seznam.cz](http://www.seznam.cz), ze zahraničních v drtivé většině uživatelé sáhnou po vyhledávači [www.google.com](http://www.google.com), který náhledy nemá.

Prostředí průzkumu bylo pro všechny respondenty uspokojující. Problém se vyskytl, když respondent zadal příliš složitý dotaz vyhledávači, systém neodpověděl žádným odkazem. Kámen úrazu se skrývá v systému Yahoo! API [4] a zvolených klíčových slovech pro plnění databáze. API dokáže na jeden dotaz odpovědět maximálně tisíci odpověďmi. Množství českých odkazů naindexovaných u Yahoo! také není veliký. Částečná kompenzace tohoto problému se vyřešila větší sadou klíčových slov pro plnění databáze odkazů. Stále ale vyhledávače neumějí reagovat na obsáhlejší české fráze. Uživatelé si ale přesto pochvalovali otázky, na které našli odpověď v českém jazyce. Našli se ale i takoví, kteří se spokojili s cenou v cizí měně.

Velkou spokojenost sklidila verze vyhledávače s náhledem. Ve většině si ji všichni respondenti pochvalovali a když byli dotázáni v které verzi výsledků se jim hledá lépe, jednoznačně odpověděli, že pro ně byly výsledky s náhledy lepší. Pouze třem respondentům bylo jedno, jaký vyhledávač použijí a ani jeden nepreferoval výsledky bez náhledů.

Náhled byl zvolen v optimální velikosti, takže byl dobře přehledný a obsah snadno rozpoznatelný. Zvlášť oceněná skutečnost byla, pokud stránka působí odpudivým vzhledem (kompletně textová, nenaformátovaná, plná reklam, chyb), lze se pomocí náhledu takovéto stránce vyhnout aniž by respondent zavítal na její obsah.

## 7. ZÁVĚR

Po zhodnocení všech výsledků a individuálních poznatků, dospívá tato práce k jednoznačnému výsledku. Náhledy mají ve vyhledávacích a internetových katalozích pozitivní vliv na rychlost hledání. Proto nelze nic jiného, než doporučit při tvorbě nového vyhledávače tento prvek zahrnout. Záleží ovšem na jeho konečném použití. Např. kategorie “homepage“, kde se prokázalo, že náhledy vliv nemají. Odkazy v této kategorii se ale většinou objevují mezi prvními na seznamu, takže jejich hledání je poněkud zjednodušeno. Potvrdilo se očekávání výsledků předchozích výzkumů [1,2] i když byly aplikovány na současné technologii.

## 7. CONCLUSION

After summary of all results and individual piece of knowledge, this work comes to the definite result. Previews in search engines have positive influence over search speed. Therefore it is recommended to involve this element while creating new search engine. Depends on final using of it. E.g. Category "homepage", where evidence, that the previews haven't have influence. Links in this category usually appears between top in the list so their searching is simplified. Also for beginners this element doesn't help them to better result. Prove true expectation of previous researchs results [1,2] even if were to be apply to current data technology.



## ÚDAJE PRO KNIHOVNICKOU DATABÁZI

Název práce	Efektivnost náhledů stránek v internetovém katalogu
Autor práce	Roman Honzák
Obor	IT
Rok obhajoby	2007
Vedoucí práce	Mgr. Miroslav Suchý
Anotace	Tato práce se zabývá problematikou náhledů stránek v internetovém vyhledávači, konkrétně však jejich efektivností při hledání.
Klíčová slova	náhledy, thumbnails, Internetové vyhledávání,

## SEZNAM POUŽITÉ LITERATURY

1. WOODRUFF, ALLISON., FAULRING, ANDREW., ROSENHOLTZ, RUTH., MORRISON, JULIE. and PIROLI, PETER. *Using Thumbnails to Search the Web*. New York : ACM Press, 2001. 198 – 205 s. ISBN:1-58113-327-8
2. DZIADOSZ, SUSAN. and RAMAN, CHANDRASEKAR. *Do Thumbnail Previews Help Users Make Better Relevance Decisions About Web Search Results?*. New York : ACM Press, 2002. 365 – 366 s. ISBN:1-58113-561-0
3. MORRISON, J. B., PIROLI, P., and CARD, S. K., A Taxonomic Analysis of What World Wide Web Activities Significantly Impact People's Decisions and Actions. Xerox PARC Tech.Report UIR-R-2000-17.
4. *Yahoo!, Developer Network, Search Web Services - Site Explorer Inbound Links API*, Copyright © 2007 [ online ] [ cit. 2007-08-10]. Dostupný z WWW:  
<http://developer.yahoo.com/search/siteexplorer/V1/inlinkData.html>
5. COLTHEART, V. (ed.). *Fleeting Memories: Cognition of Brief Visual Stimuli*. MIT Press: Cambridge, MA, 1999 (pp. 32-70).
6. *Lupa: Server o českém internetu, Suché komentáře, Náhledy stránek: Kdy ano a kdy ne*, 24. srpen, 2006 [online ]. Suchý, Miroslav. [cit. 2007-08-10]. Dostupný z WWW:  
<http://suchy.blog.lupa.cz/0608/nahledy-stranek-kdy-ano-a-kdy-ne>.