

**Univerzita Karlova v Praze**

**Filozofická fakulta**

**Ústav informačních studií a knihovnictví**

Studijní program: informační studia a knihovnictví

Studijní obor: informační studia a knihovnictví

# **Diplomová práce**

Bc. Jana Pačísková

## **Aspekty a trendy inteligentního vyhledávání**

Intelligent information retrieval and its trends

**Prohlášení:**

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 21. dubna 2014

.....  
podpis studenta

### **Identifikační záznam**

PAČÍSKOVÁ, Jana. *Aspekty a trendy inteligentního vyhledávání = Intelligent information retrieval and its trends*. Praha, 2014. 78 s., 8 s. příl. Diplomová práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí práce Richard Papík.

### **Abstrakt (česky)**

Diplomová práce se zabývá vyhledáváním informací v kontextu jeho historického vývoje a představuje trendy v integraci inteligentních prvků a tím vznik tzv. inteligentního vyhledávání. Jednotlivé inteligentní prvky jsou popsány v samostatné kapitole, v následující kapitole je pak představeno jejich použití včetně konkrétních příkladů. Práce také mapuje zpracování tématu inteligentního vyhledávání ve vybraných institucích v Česku i v zahraničí; výsledky tohoto průzkumu pro české prostředí jsou prezentovány v příložené rešerši.

**Klíčová slova (česky):** vyhledávání informací, inteligentní vyhledávání informací, umělá inteligence, internet

**Abstrakt (anglicky):**

This thesis is focused on information retrieval in the context of its historical development, it presents trends in integration of intelligent features in it, and thus the emergence of intelligent information retrieval. Individual intelligent elements are described in a separate chapter, following chapter then introduces their use, including specific examples. Thesis also traces research on the topic of intelligent information retrieval in selected institutions both in the Czech republic and abroad; results of this survey for Czech republic are presented in the enclosed search.

**Klíčová slova (anglicky):** information retrieval, intelligent information retrieval, artificial intelligence, internet

# OBSAH

<b>SEZNAM ZKRATEK.....</b>	<b>7</b>
<b>PŘEDMLUVA .....</b>	<b>9</b>
<b>ÚVOD .....</b>	<b>11</b>
<b>1 VYHLEDÁVÁNÍ INFORMACÍ .....</b>	<b>12</b>
1.1 Definice .....	12
1.2 Rešeršní systém.....	12
1.3 Typologie.....	15
1.4 Historie .....	16
1.4.1 Generace vyhledávacích strojů .....	17
1.4.2 Inteligentní vyhledávání .....	18
1.5 Modely.....	18
1.5.1 Booleovský model vyhledávání .....	19
1.5.2 Fuzzy model vyhledávání .....	20
1.5.3 Vektorový model vyhledávání .....	20
1.5.4 Pravděpodobnostní model vyhledávání.....	21
1.6 Webové vyhledávání .....	21
1.6.1 Internetové vyhledávače (web search engines).....	23
<b>2 INTELIGENTNÍ VYHLEDÁVÁNÍ.....</b>	<b>25</b>
2.1 Vymezení pojmu .....	25
2.2 Umělá inteligence.....	26
2.2.1 Dějiny umělé inteligence .....	27
2.2.2 Artificial general intelligence.....	28
2.3 Inteligentní prvky ve vyhledávání informací.....	29
2.3.1 HCI a uživatelské rozhraní.....	30
2.3.2 „Inteligentní chování“ uživatelského rozhraní.....	31
2.3.3 Počítačové vidění .....	32
2.3.4 Počítačový sluch.....	33
2.3.5 Počítačový hmat .....	34
2.3.6 Strojové učení.....	35
2.3.7 Zpracování přirozeného jazyka .....	37
2.4 Terminologická analýza .....	38
<b>3 POUŽITÍ INTELIGENTNÍHO VYHLEDÁVÁNÍ.....</b>	<b>41</b>
3.1 Vyhledávání obrazových informací .....	41
3.1.1 Vyhledávání obrázků kreslením.....	43
3.1.2 Vyhledávání tváří .....	45
3.1.3 Vyhledávání chemických vzorců .....	47

3.2	Vyhledávání hudebních informací .....	49
3.3	Vyhledávání audiovizuálních informací.....	52
3.4	Enterprise search.....	53
3.4.1	Nalezitelnost.....	54
3.4.2	Enterprise search versus webové vyhledávání.....	56
<b>4</b>	<b>SMĚRY VÝVOJE VYHLEDÁVÁNÍ INFORMACÍ V UZAVŘENÝCH INFORMAČNÍCH SYSTÉMECH A NA VOLNÉM INTERNETU .....</b>	<b>58</b>
4.1	Uživatelské rozhraní a nápověda .....	58
4.2	Rozšíření služeb .....	60
<b>5</b>	<b>INTELIGENTNÍ VYHLEDÁVÁNÍ JAKO PŘEDMĚT VÝZKUMU .....</b>	<b>61</b>
5.1	Odborná pracoviště a univerzity v ČR .....	61
5.2	Zahraniční odborná pracoviště.....	67
	<b>ZÁVĚR .....</b>	<b>70</b>
	<b>SEZNAM POUŽITÉ LITERATURY .....</b>	<b>72</b>
	<b>SEZNAM OBRÁZKŮ .....</b>	<b>79</b>
	<b>PŘÍLOHA Č. 1: REŠERŠE VE VYBRANÝCH ČESKÝCH INSTITUCÍCH NA TÉMA INTELIGENTNÍ VYHLEDÁVÁNÍ.....</b>	<b>0</b>

## SEZNAM ZKRATEK

AGI	artificial general intelligence (umělé bytí, umělý život)
apod.	a podobně
atd.	a tak dále
CIIR	Center for Intelligent Information Retrieval
cit.	citováno, citace
ČVUT	České vysoké učení technické
ERP	enterprise resource planning
GSA	Google Search Appliance
HCI	human-computer interaction (komunikace člověk-počítač)
HTML	hypertext markup language (hypertextový značkovací jazyk)
HTTP	hypertext transfer protocol
idf	inverse document frequency (převrácená četnost slova ve všech dokumentech)
IEEE	Institute of Electrical and Electronics Engineers
IIR	intelligent information retrieval (inteligentní vyhledávání informací)
IR	information retrieval (vyhledávání informací)
KISK	Kabinet informačních studií a knihovnictví (Masarykova univerzita v Brně)
MFF	Matematicko-fyzikální fakulta (Univerzita Karlova)
mj.	mimo jiné
např.	například
NLP	natural language processing (zpracování přirozeného jazyka)
resp.	respektive
SIGIR	Special Interest Group on Information Retrieval
TDKIV	Terminologická databáze knihovnictví a informační vědy
tj.	to jest

tf	term frequency (četnost slova v dokumentu)
tzv.	takzvaný
UI	umělá inteligence
ÚISK	Ústav informačních studií a knihovnictví (Univerzita Karlova)
UK	Univerzita Karlova
URL	uniform resource locator (unifikovaný lokátor zdroje)
VŠE	Vysoká škola ekonomická
WSD	word sense disambiguation (rozlišení významů slov)
XML	extensible markup language (rozšiřitelný značkovací jazyk)



## PŘEDMLUVA

Ve své práci se zabývám inteligentním vyhledáváním jako vědní disciplínou i jako jednou z aplikačních oblastí umělé inteligence. Téma jsem si zvolila ze dvou hlavních důvodů. Zaprvé proto, že mě problematika vyhledávání informací zajímá již od bakalářského studia – ve své bakalářské práci jsem se zabývala vyhledávacím modulem knihovního systému Aleph. A druhým důvodem je vlastně sám obor umělé inteligence, jehož cíle a vize mi připadají veskrze fascinující a vidím v nich ohromný potenciál uplatnění v mnoha odvětvích lidské činnosti.

Poprvé jsem se s koncepcí umělé inteligence a její souvislostí s naším oborem setkala hned v prvním ročníku bakalářského studia v předmětu *Základy informační vědy*, kde jsme byli seznámeni s existencí neuronových sítí a expertních systémů. Problematika mě zaujala a během studia jsem se snažila účastnit se výběrových přednášek, které by probíhaly na podobná nebo příbuzná témata. Velkým přínosem v tomto ohledu pro mě bylo navštěvování přednášek na téma *Informatika a kognitivní vědy* na MFF UK, v rámci nichž jsem se zúčastnila experimentu realizovaného na Fyziologickém ústavu Akademie věd ČR, který zkoumal orientaci člověka v prostoru virtuální reality.

V úvodní kapitole je pro lepší návaznost a v souladu se zadáním práce stručně představena historie vyhledávání informací obecně, modely vyhledávání a generace vyhledávacích strojů. Následuje kapitola o inteligentním vyhledávání, ve které je představena terminologie související s touto oblastí a kde rovněž rozebírám jednotlivé prvky, kterými se inteligentní vyhledávání vyznačuje. V další kapitole potom uvádím konkrétní příklady aplikace inteligentního vyhledávání včetně některých příkladů a názorných ukázek.

Práce také obsahuje kapitolu o zpracování tohoto tématu v oblasti výzkumu a vývoje především na vysokých školách a ve výzkumných institucích. Tato kapitola je doplněna rešerší představující konkrétní výstupy v této oblasti v České republice. Rešerše byla provedena v repozitářích a knihovních katalozích jmenovaných institucí, případně online průzkumem jejich webových stránek, a nachází se v příloze ([příloha č. 1](#)).

Cílem a přínosem práce je především vytvoření přehledu problematiky inteligentního vyhledávání, který může sloužit jako podklad pro následnou

výzkumnou a studijní činnost, shrnutí současných trendů a nastínění směrů vývoje v této oblasti, které by se mohly stát předmětem dalšího zkoumání.

Největším problémem, se kterým jsem se během psaní práce potýkala, byla širší tématu. Problematika vyhledávání informací je velmi obsáhlá. Snažila jsem se práci koncipovat tak, aby sloužila jako přehledová studie a zmiňovala se o nejvýznamnějších aspektech inteligentního vyhledávání, byť třeba jen stručně.

Použitá literatura obsahuje z podstatné části elektronické zdroje, což je dáno zaměřením práce na aktuální trendy a obecně povahou tématu, které je relativně mladé a dosud nebylo komplexně zpracováno. Při psaní jsem kromě citované literatury využívala též vlastních znalostí, které jsem z oblasti inteligentního vyhledávání dosud nasbírala. Soupis použité literatury je tvořen dle ISO 690, v textu práce používám harvardský citační styl. Rozsah práce je 79 stran a 8 stran přílohy.

## **Poděkování**

Na tomto místě bych ráda poděkovala vedoucímu diplomové práce doc. PhDr. Richardu Papíkovi, Ph.D. za konzultace, motivaci, podporu, cenné rady a veškerou další pomoc při psaní této práce.

Dále bych chtěla poděkovat své rodině a přátelům za trpělivost, shovívavost a podporu ve chvílích nejtěžších.

# ÚVOD

„*Nic není v rozumu, co by nebylo dříve ve smyslech.*“ Francis Bacon

I když se to na první pohled nemusí zdát zřejmé, vyhledávání informací je činnost, které je od pradávna součástí každodenního života člověka. V nejobecnějším měřítku se jedná o získání informace, kterou zatím nemáme, ale k něčemu ji potřebujeme. I ranní pohled z okna, abychom věděli, jaké je počasí a co si tedy obléknout, je ve své podstatě určitou formou vyhledávání informací.

Pokud svůj pohled na problematiku zúžíme a vezmeme vyhledávání informací jako činnost pro podporu práce a rozhodování čerpáním z informačních zdrojů, dlouhou dobu bylo doménou vzdělanců a knihovníků, kteří ke zkvalitnění vyhledávání informací přispívali svými vlastními znalostmi. Pomáhali například tím, že vhodným vyptáváním donutili člověka formulovat přesně svou informační potřebu, nebo přispívali svými zkušenostmi a pamětí, díky nimž věděli, kde přesně by se daná informace dala nalézt. Velká proměna vyhledávání informací nastala s rozvojem informačních technologií, kdy se k vyhledávání začaly používat stroje. S nástupem internetu pak došlo k jakési demokratizaci vyhledávání informací – tato činnost se náhle stala přístupná všem, kdo o ni měli zájem.

S narůstajícím objemem lidského poznání a masovou využívaností vyhledávání na internetu je stále potřeba zlepšovat podmínky a zkvalitňovat výkon vyhledávacích strojů. A tak současným trendem je vlastně do jisté míry návrat k původnímu konceptu, kdy k procesu vyhledávání informací přispívali tehdejší odborníci svými znalostmi. Snažíme se, aby podobným způsobem dokázaly podpořit vyhledávací proces i dnešní stroje. Tedy aby podobně jako lidé dokázali sbírat zkušenosti, učit se, komunikovat s uživatelem způsobem, který je pro něj přirozený, poradit mu, nebo i použít empatii a *cítit*, či dokonce *souznít* s ním. Dáváme tedy strojům smysly a zároveň doufáme, že časem jim budeme moci dát i rozum.

# 1 Vyhledávání informací

## 1.1 Definice

Definice termínu **vyhledávání informací** existuje velké množství. Záleží vždy na tom, z jakého úhlu a v jakém kontextu je na problematiku nahlíženo. Uvedu definici z Knihovnické terminologické databáze (TDKIV), protože je dostatečně obecná, nicméně nahlíží na problematiku z hlediska našeho oboru a připadá mi tedy pro účely této práce vhodná. Říká, že **vyhledávání informací (selekce informací)** je *„činnost, jejímž cílem je identifikace relevantních dokumentů nebo informací v informačních zdrojích (např. plnotextové nebo bibliografické databáze). Vyhledávání informací probíhá obvykle na základě konkrétního požadavku uživatele za pomoci dotazovacích a selekčních jazyků“* [ŠVEJDA, 2003a]. Pozornost bude v této práci věnována pouze elektronickému vyhledávání informací a to v jeho (jak říká definice) „obvyklé“ podobě, tedy dotazově orientovanému vyhledávání.

## 1.2 Rešeršní systém

Dalším důležitým pojmem je **rešeršní systém**, jehož funkcí je právě *„vyhledávání a výběr (selekce) informací relevantních informačním požadavkům uživatelů“* [ŠVEJDA, 2003]. Chowdhury jej označuje za jakýsi most mezi světem tvůrců a generátorů informací na jedné straně, a světem uživatelů těchto informací na straně druhé [CHOWDHURY, 2004, s. 3]

Definice nám sice říká, co je cílem vyhledávání informací, ovšem jak toho dosáhnout, o tom už se nezmiňuje. V průběhu vývoje, s proměnami technologií, společnosti, uživatelského chování apod., vyvstala celá řada problémů – zpracování přirozeného jazyka, přesné definování informační potřeby, formulace informační potřeby do podoby informačního dotazu, identifikace relevantních dokumentů atd.

Ukázalo se, že **identifikace relevantních dokumentů** je jedním z klíčových aspektů, a že dosáhnout uspokojivých výsledků v této oblasti je nesmírně obtížné, a to i když se soustředíme pouze na vyřešení problému s relevancí a nebereme ohled na další aspekty vyhledávání informací, jako je jeho rychlost nebo uživatelská přístupnost. Jedním z hlavních cílů rešeršního systému je právě nalezení co nejvíce

relevantních dokumentů rešeršnímu dotazu, a zároveň vyloučení těch, které relevantní nejsou [BAEZA-YATES, 1999, s. 2]. Problém ovšem tkví v tom, že relevance dokumentů je relativní – dokument může být tematicky relevantní, ovšem jeho uživatelská relevance může být klidně nulová, např. ve chvíli, kdy uživatel daný dokument již četl, je neaktuální, nebo je v jazyce, kterému nerozumí apod. [CROFT, 2010, s. 5].

Důležitým krokem k výkonnému rešeršnímu systému je stanovit si požadavky, které chceme, aby naplnil. Sklenák [2001, s. 225-226] uvádí celkem deset charakteristik ideálního rešeršního systému:

- *minimalizace doby odezvy mezi zadáním dotazu uživatele a odpovědí systému,*
- *uchovávání užitečných údajů,*
- *účelné uživatelské rozhraní*
- *souběžná práce více uživatelů s týmiž daty (resp. dokumenty),*
- *trvalost uchovávání a minimalizace redundance dat,*
- *dotazovací jazyk blízký přirozenému jazyku,*
- *možnost zabezpečeného přístupu,*
- *co nejjednodušší vyhledávání s minimálními nároky na učení uživatele,*
- *efektivní přidávání, rušení a aktualizace dat (resp. dokumentů),*
- *trvalá dostupnost bez jakýchkoliv časových omezení.*

Sklenák se také věnuje požadavkům uživatelů na vyhledávání informací, které shrnuje následovně [SKLENÁK, 2001, s. 226-227]:

- ***zpětná vazba relevance*** – *uživatel ve vyhledané množině dokumentů označí pro něj nejzajímavější dokumenty, a vyhledávací systém pak na základě těchto dokumentů automaticky vytvoří nový dotaz, pomocí kterého budou vyhledány podobné dokumenty*
- ***extrahování informací*** – *proces identifikace databázových entit, jejich atributů a vztahů v plném textu (např. identifikace firem, jejich aktivit, nových služeb a výrobků apod.)*

- **multimediální vyhledávání** – požadavek na multimediální indexování, aniž by různé mediální formy (obrázky, zvuky a videosekvence) musely být doplňovány konvenčním textovým popisem (deskriptory)
- **efektivní vyhledávání** – tato oblast tvoří jádro výzkumu v oblasti IR již více než 30 let; nejčastěji používanými mírami (i při hodnocení kvality různých systémů) jsou úplnost a přesnost – v první řadě jde především o jejich zvyšování
- **filtrování informací** – proces identifikace relevantních informací v toku nových informací; zatímco u „klasického“ vyhledávání je jeden dotaz konfrontován s velkým objemem uložených dokumentů, při filtrování je porovnávána sada definovaných profilů s jednotlivými dokumenty; pokud je dokument relevantní, je zaslán uživateli, jemuž přísluší daný profil (profil reprezentuje trvalou informační potřebu, jeho formulace bývá komplexnější než u jednorázových dotazů); zkoumají se možnosti využití tzv. informačních agentů
- **„jazyková magie“** – jednou z nejčastějších příčin neúspěchu při vyhledávání je volba jiných slov než v databázovém indexu (slovníku), tzn. informační potřeba je popsána jinými slovy, než která jsou použita v relevantních dokumentech; problém se řeší např. vhodnou, automatickou expanzí dotazu, používají se různé techniky, např. Latent Semantic Indexing (skryté sémantické indexování) nebo použití zabudovaného tezauru
- **flexibilní a efektivní indexování** – jeden z nejdůležitějších předpokladů efektivního vyhledávání; kritickými metrikami jsou především rychlost odezvy na dotaz a rychlost indexování; zajímavým aspektem indexování je nutnost vypořádání se s bohatostí formátů dokumentů
- **distribuované vyhledávání** – prostřednictvím počítačových sítí mají uživatelé k dispozici mnoho informačních zdrojů, problém spočívá v jejich lokalizaci, jejich optimálním výběru, provedení simultánního dotazu, zatřídění výsledků podle normované škály relevance (může být

*nekompatibilní s individuálními koeficienty relevance konkrétních systémů)*

Těžko lze nalézt úplné řešení všech problémů a přijatelné odpovědi na zmíněné požadavky najednou, navíc při neustále se měnících podmínkách a objevování nových výzev v oblasti vyhledávání informací. Přesto se řešení hledají, i kdyby jen částečná nebo pouze pro jednotlivé problémy. Jedním z „produktů“ hledání takových řešení je i tzv. inteligentní vyhledávání.

### 1.3 Typologie

Existuje mnoho typologií vyhledávání podle různých aspektů. Uvádět je všechny je, myslím, zbytečné, navíc je jich tolik, že by to ani nebylo možné – podle míry strukturovanosti dat, druhů prohledávaných informací nebo jejich formy, podle způsobu konstrukce dotazu atd. Uvedu pouze rozdělení, která jsou pro potřeby práce důležitá a měla vliv na vývoj inteligentního vyhledávání.

V elektronickém prostředí lze na vyhledávání informací nahlížet ze dvou různých úhlů [BAEZA-YATES, 1999, s. 7]:

1. **Computer-centered** (zaměření na stroj, počítač)
2. **Human-centred** (zaměření na člověka, uživatele)

V prvním jde o technickou stránku vyhledávacího procesu – způsob sestavování indexu, výkonnost a efektivitu systému, rychlost reakce na dotaz atd., zatímco v druhém případě je pozornost věnována uživatelskému chování, porozumění potřebám uživatele a tomu, jak by tyto aspekty měly ovlivnit organizaci informací a způsob vyhledávání [BAEZA-YATES, 1999, s. 7].

S neustále se zvětšujícím objemem přístupných informací díky rozvoji internetu lze vyhledávání dělit podle toho, v jak hutném informačním prostředí, co se týče množství prohledávaných dokumentů, musí operovat [MANNING, 2008, s. 2]:

1. **Web search** – pracuje s ohromným množstvím dokumentů a navíc se musí přizpůsobit prostředí webu
2. **Enterprise, institutional, domain-specific search** – typicky pracuje v centralizovaném úložišti se specifikovanými přístupy

### 3. *Personal information retrieval* – vyhledávání na osobním počítači, v e-mailu apod.

Vyhledávání lze také dělit podle způsobu, jak reaguje na uživatelův dotaz. Vyhledávání typu „push“ je takové, které uživatel jednou iniciuje, nastaví mu parametry a ono pak probíhá po stanovenou dobu v určitých intervalech automaticky a výsledky prezentuje uživateli. Naproti tomu u typu „pull“ každé sadě výsledků předchází podnět od uživatele, je tedy vždy iniciováno znovu [MANNING, 2008, s. 291].

## 1.4 Historie

Když se na přelomu padesátých a šedesátých let k vyhledávání informací začalo využívat počítačů, objevily se dvě základní otázky – jak indexovat dokumenty a jak je poté nalézt. To byla chvíle, kdy se zrodilo vyhledávání informací jakožto výzkumná disciplína, která se bude snažit na tyto otázky nalézt odpovědi [CROFT, 2012].

V šedesátých letech se v oblasti vyhledávání přešlo od otázky „Jak využít počítač k vyhledávání?“ k hledání způsobů, jak vyhledávání na počítačích zlepšit. Začaly se používat **řadící algoritmy** k pořadání výsledků, přičemž se pochopitelně narazilo na problém s posuzováním relevance – částečné řešení poskytla **relevantní zpětná vazba** (uživatel označí ve výsledcích vyhledávání ty dokumenty, které jsou k jeho dotazu skutečně relevantní), která byla prvním využitím strojového učení ve vyhledávání informací [CROFT, 2012]. Začalo se také využívat **shlukování** dokumentů podobného obsahu, které napomohlo ke zvýšení počtu dokumentů odpovídajících rešeršnímu dotazu.

V podstatě všechny vyhledávací programy v té době používaly **Booleovské** vyhledávání, které je dodnes hojně používané. Nicméně objevily se i jiné modely vyhledávání – v sedmdesátých letech to bylo **vektorové** a **pravděpodobnostní** vyhledávání. V té době se také zrodil nápad na váženou indexaci, tehdy konkrétně vážení založené na **četnosti termínu v dokumentu** (tf), brzy nato i na **převrácené četnosti termínu ve všech dokumentech** (idf) a konečně na jejich kombinaci: **tf-idf**. Dalším důležitým krokem bylo odhalení **závislosti** slov, totiž nápad, že některá slova jsou spolu svázána více než jiná a jsou tedy na sobě do jisté míry závislá. Tato



myšlenka byla dále rozvíjena a využita při posuzování relevance pro řazení výsledků [CROFT, 2012].

V dalších letech došlo k obohacování již známých modelů vyhledávání o další možnosti, začaly se do nich zapracovávat dříve objevené možnosti posuzování relevance, vylepšovaly se algoritmy používané při vyhledávání (např. pro stemming) atd.

Devadesátá léta se nesou ve znamení **webu**. Se vznikem internetu a s ním spojeným prudkým nárůstem množství informací vznikl silný tlak na vytvoření vyhledávacích strojů, které by dokázaly efektivně fungovat v tak odlišném prostředí.

### **1.4.1 Generace vyhledávacích strojů**

Vývoj vyhledávacích strojů, které už nebyly jen teoriemi, nápady nebo experimenty, ale skutečně se používaly v praxi, lze v elektronickém prostředí rozdělit na tři generace [BAEZA-YATES, 1999, s. 7].

První generace byla v podstatě prostou automatizací předchozích vyhledávacích technik. Protože knihovny byly jedněmi z prvních institucí, které implementovaly rešeršní systémy pro vyhledávání informací, byla první generace vlastně jen jiným druhem lístkového katalogu a umožňovala tedy vyhledávání podle autora a názvu, ale nic víc.

S objevem potenciálu automatizovaného vyhledávání se původní možnosti rozšířily – např. o vyhledávání podle předmětových hesel nebo klíčových slov a o některé možnosti pro vytváření komplexnějších rešeršních dotazů. Zrodila se druhá generace vyhledávacích strojů.

Třetí generace, do které spadají i dnešní rešeršní systémy, se začala vyvíjet díky pokroku v informačních technologiích, které umožnily zlepšení funkcionality rešeršních systémů i jinými způsoby než možnostmi vyhledávání – zvýšením kvality grafického rozhraní, integrací hypertextu, implementací otevřené architektury systému atd.

Další významný posun zaznamenala oblast vyhledávání s rozvojem webu, kdy vyhledávání najednou nebylo už jen záležitostí specialistů a vědců, komunita uživatelů se mnohonásobně rozrostla.

Vyhledávání informací stále řeší velké množství problémů, nicméně jsou i takové, v nichž se za léta vývoje zaznamenal značný posun. Úplně první stroj použitý k vyhledávání informací koncem čtyřicátých let dvacátého století dokázal zpracovat 120 slov za minutu [CROFT, 2012]. Dnes například internetové vyhledávače pracují s nesrovnatelně větším množstvím informací a vyhledávací časy se přitom pohybují v řádu desetin sekund.

### **1.4.2 Inteligentní vyhledávání**

Tuto podkapitolku zařazuji proto, abych uvedla do historických souvislostí inteligentní vyhledávání. Tato návaznost je totiž poněkud problematická, záleží totiž na tom, v čem právě ta „intelligence“ spočívá, na což se názory různí. Podrobnější vysvětlení a nastínění současného stavu se objeví v samostatné kapitole.

Nejčastěji evokuje inteligentní vyhledávání procesy spojené s umělou inteligencí. Pokud hledáme historické začátky, je to asi pohled správný. Obor umělé inteligence se začal rozvíjet v 50. a 60. letech. Začlenění některých jeho odvětví do oblasti vyhledávání přišlo v podstatě záhy s využitím strojového učení. Postupně s objevujícími se problémy s posuzováním relevance a zjištěnými nevýhodami vyhledávacích modelů se zdálo začlenění umělé inteligence výhodným. V 70. letech vznikají první expertní systémy, v 80. letech už se běžně mluví o inteligentních rešeršních systémech [JONES, 1983] a využití neuronových sítí ve vyhledávání informací [BELEW, 1989].

## **1.5 Modely**

Pro vypořádání se s nejrůznějšími problémy, na které vyhledávání informací naráželo, se začaly vytvářet modely.

Klasické modely se dělí na tři skupiny, které vycházejí z historických kořenů vyhledávání: **booleovské**, **vektorové** a **pravděpodobnostní** modely. Postupem času se v každé skupině začaly vyvíjet alternativní verze, ovšem založené na stejném základu. Booleovské modely byly obecně nazvány **množinovými** (set-theoretic), protože jejich základem je reprezentace dokumentů a dotazů množinami termínů. Vektorovým modelům se říká **algebraické** (algebraic), protože funkci reprezentace dokumentů a dotazů namísto množin plní vektory v n-rozměrném prostoru.

**Pravděpodobnostní** (probabilistic) modely jsou založeny na teorii pravděpodobnosti a byly podle ní také pojmenovány [BAEZA-YATES, 1999, s. 20-21].

### **1.5.1 Booleovský model vyhledávání**

Booleovský model nese svůj název po Georgi Booleovi, britském matematikovi a logikovi, který začal používat ve svém symbolickém logickém systému tři operace pro kombinace jednotlivých tvrzení: logickou konjunkci (AND), disjunkci (OR) a negaci (NOT). Díky jednoduchosti této koncepce se Booleovský model vyhledávání začal používat masově v různých rešeršních systémech včetně online katalogů a databází.

Ovšem jako každý model měl i ten Booleovský své slabiny, které se záhy po jeho implementaci začaly ukazovat. První problém spočíval v obtížnosti transformace informační potřeby do správné podoby informačního dotazu. Celkově je vytváření kombinací termínů v dotazu pomocí uvedených operátorů problematické, obzvláště při jejich větším počtu – záběr dotazu se pak ukázal buď jako příliš široký, nebo naopak příliš úzký. Z toho vyplývala další potíž, totiž že při širokém záběru dotazu systém vygeneruje obrovský počet výsledků a je v podstatě nemožné přesně identifikovat ten, který by odpovídal původnímu požadavku (obdobně u příliš úzkého dotazu, kdy jsou mnohé relevantní dokumenty opomenuty).

Třetím problémem je nemožnost řazení výsledků podle relevance, protože mechanismus vyhledávání spočívá v prostém ověření přítomnosti dotazovaných termínů v daném záznamu. Posouzení relevance je tedy binární, buď-anebo, ale nic mezi tím, z čehož vyplývá, že systém musí nutně považovat všechny výsledky za stejně relevantní, tj. že neumožňuje zahrnout do výsledků dokumenty, které dotazu odpovídají jen částečně.

**Rozšířená verze Booleovského modelu** představená v roce 1983 Gerardem Saltonem, Edward A. Foxem, and Harrym Wuem se pokouší řešit právě posledně zmiňovaný problém. Přidala do Booleovského modelu vážení termínů a možnost částečné shody s dotazem, které si „vypůjčila“ z vektorového modelu (podrobnější popis v příslušné podkapitole o vektorovém modelu vyhledávání). Rozšířený Boole je tedy takovým hybridním modelem, protože má charakteristiky jak množinového, tak vektorového vyhledávání [BAEZA-YATES, 1999, s. 41].

### **1.5.2 Fuzzy model vyhledávání**

Tento množinový model je založen na **fuzzy logice**, ve které se výroky ohodnocují mírou pravdivosti. Na rozdíl od klasické logiky, kde se používá logických hodnot nula nebo jedna, operuje fuzzy logika navíc se všemi hodnotami, které se nacházejí mezi tím.

Nejjednodušší je asi vysvětlit princip fuzzy modelu na příkladu. Řekněme, že máme odbornou knihu, která se zabývá chemickými látkami, jejich složením a vlastnostmi, například jedy. V některých částech knihy se hovoří také o využití jedů v medicíně. Kniha se tedy zabývá lékařskou tematikou jen částečně, přesto je to důležité. Pro téma lékařství se tedy přiřadí hodnota pro „*stupeň začlenění prvku do množiny*“ [GROSSMAN, 2004, s. 85], což je vlastně funkce nabývající hodnot z intervalu  $<0; 1>$ . Obdobně se ohodnotí i ostatní prvky, čímž vznikne fuzzy množina popisující daný dokument.

S fuzzy množinami se pak pracuje pomocí množinových operací. Pro průnik dvou (nebo více) fuzzy množin se bere minimum dvou (nebo více) funkcí pro stupeň začlenění pro jeden a ten samý prvek v těchto množinách, sjednocení množin naopak maximum funkcí [GROSSMAN, 2004, s. 85]. Byly představeny i modely, které nevyužívají hodnoty maxima a minima funkce, ale pracují s algebraickými operacemi (součet, součin) [BAEZA-YATES, 1999, s. 37-38].

### **1.5.3 Vektorový model vyhledávání**

Vektorový model počítá míru podobnosti pomocí reprezentace dokumentů a dotazů vektory [SALTON, 1975]. Přesněji řečeno reprezentace termínů obsažených v dokumentu, a termínů obsažených v dotazu. Tato hodnota se nazývá **koeficient podobnosti** a počítá se nejčastěji na základě úhlu, který tyto vektory svírají.

V tomto modelu se navíc uplatňuje nebinární vážení termínů, které je realizováno buď manuálně uživatelem, nebo automaticky systémem na základě frekvence výskytu termínu v celé sbírce dokumentů [GROSSMAN, 2004, s. 12].

Právě vážení termínů a počítání podobnosti, které umožňuje vyhledání dokumentů odpovídajících dotazu jen částečně, jsou považovány za největší výhody vektorového modelu. Nevýhodou může být skutečnost, že termíny jsou brány jako na sobě nezávislé. Baeza-Yates ale naznačuje, že se možná o slabinu ani nejedná, protože závislost termínů je často platná jen v některých oblastech a aplikovat ji na

celou sbírku dokumentů by mohlo celkovému výkonu modelu spíše ublížit [BAEZA-YATES, 1999, s. 30].

#### **1.5.4 Pravděpodobnostní model vyhledávání**

Tento model využívá podobně jako model vektorový koeficient podobnosti, ovšem počítá ho na základě pravděpodobnosti, že daný dokument bude relevantní dotazu.

V prvním modelu, který představili v roce 1960 Maron a Kuhns, jde o pravděpodobnost, s jakou ohodnotí uživatel, který přijde s určitým konkrétním dotazem, libovolný dokument jako relevantní. Tedy jde o to, kolik uživatelů se stejným dotazem označí které dokumenty za relevantní [CHOWDHURY, 2004, s. 174-175].

Jiný přístup zvolili v roce 1976 S. E. Robertson a K. Spärck Jones. V jejich modelu jde v podstatě o myšlenku, že ke každému dotazu existuje ideální množina výsledků, jejíž přesné vlastnosti nejsou v době zadávání dotazu známy. Na proces vyhledávání lze tedy nahlížet jako na „*specifikaci vlastností ideální množiny výsledků*“ [BAEZA-YATES, 1999, s. 31], které ovšem při zadání dotazu nemohou být než odhadnuty (na základě charakteristik termínů z dotazu). Poté je vygenerována předběžná množina výsledků, která je předložena uživateli kvůli relevantní zpětné vazbě. Na základě této zpětné vazby následně dojde k upřesnění vlastností oné ideální množiny výsledků, přičemž opakováním tohoto procesu se má docílit co největšímu přiblížení se onomu myšlenému ideálu [ROBERTSON, 1976].

Výhodou modelu je řazení výsledků podle pravděpodobnosti, že jsou k dotazu relevantní. Za nevýhody pak lze pokládat nutnost *odhadnout* rozdělení na relevantní a nerelevantní dokumenty a skutečnost, že není brána v potaz frekvence výskytu termínů v dokumentu [BAEZA-YATES, 1999, s. 34].

## **1.6 Webové vyhledávání**

Historie webu sahá až do roku 1945, tedy alespoň v teorii, nicméně opravdový boom zaznamenal až v 90. letech, kdy Tim Berners-Lee vytvořil World Wide Web v takové podobě, v jaké ho známe dnes [LEVENE, 2010, s. 3-4]. V dnešní době se jedná o velmi důležitou oblast vyhledávání informací, i kdyby jen proto, že je masově

využívaná. Pokroky v oblasti webového vyhledávání probíhají v podstatě neustále, což je dáno také proměnami technologií – přístup k internetu už lze získat nejen prostřednictvím počítačů, ale také přes nejrůznější mobilní zařízení.

Web funguje na základě tří konvencí [MANNING, 2008, s. 386]:

- URL (uniform resource locator), které slouží k identifikaci webových stránek
- HTTP (hypertext transfer protocol) sloužící k přenosu dat mezi serverem a prohlížečem
- HTML (hypertext markup language) sloužící k zobrazení webových stránek

Prostředí webu má svá specifika, kterým se vyhledávání musí nutně přizpůsobit. V literatuře jsou nazývána různě [CHOWDHURY, 2004, s. 330-332; MANNING, 2008, s. 387-388; BAEZA-YATES, 1999, s. 367-371], nicméně ve výsledku jde vždy o totéž. Shrnout je lze následovně:

1. **Množství informací** – ve srovnání s klasickým vyhledáváním informací v minulosti se musí webové vyhledávače vyrovnat s naprosto nesrovnatelně větším objemem informací, které je potřeba prohledávat, nemluvě o faktu, že jich neustále velkou rychlostí přibývá a mnoho z nich je redundantních nebo nekvalitních.
2. **Otevřenost** – web je otevřená síť. Každý uživatel má možnost vytvořit vlastní webové stránky, jejich množství není nijak omezeno nebo regulováno. S tím souvisí následné obtížné hodnocení relevance dokumentů – v klasickém vyhledávání informací se posuzuje relevance dokumentu vůči rešeršnímu dotazu. Na webu to ale nestačí, jelikož obsah stránek může být nepravdivý nebo zavádějící, takže je potřeba vzít v úvahu i další faktory a nějakým způsobem posoudit současně *důvěryhodnost* informace.
3. **Proměnlivost** – jde ruku v ruce s otevřeností webu. Obsah stránek se mění, rychle vznikají i zanikají, nebo naopak zůstávají léta naprosto nedotčeny, mohou měnit umístění apod.

4. **Různorodost** – web obsahuje informace nejrůznějších druhů (text, obrázky, videa...), navíc ještě v různých formátech a jazycích i alfabetach.
5. **Distribuovanost** – data jsou uložena na mnoha serverech po celém světě a není používán žádný jednotný standard pro vytváření a zpracování webových informačních zdrojů.

Tyto charakteristiky webu neustále tlačí na vývojáře vyhledávacích strojů, aby kontrolovaly efektivitu vyhledávačů a případně je vylepšovali. Tento tlak se promítá i do oblasti vyhledávání informací obecně a pozitivně ovlivňuje jeho výzkum.

Už bylo zmíněno, že web obsahuje ohromné množství informací. Ne všechny jsou ovšem otevřeně přístupné – mnoho z nich je skryto v části webu, do které se klasické vyhledávače nedostanou. Lze tedy říct, že web se vlastně skládá ze dvou vrstev: z **povrchového webu**, který je indexován běžnými vyhledávači a je tedy přístupný všem, a z **hlubokého webu**, který lze obvykle prohledávat pouze z konkrétních přístupových míst a vyžaduje autentikaci uživatele nebo spuštění příslušného programu [CHOWDHURY, 2004, s. 331]. Typickým představitelem informačních zdrojů hlubokého webu jsou všemožné komerční databáze a databázová centra [PAPÍK, 2011, s. 68].

### **1.6.1 Internetové vyhledávače (web search engines)**

Vyhledávání v prostředí povrchového webu je realizováno pomocí vyhledávačů. Kvůli distribuované povaze webu v podstatě není možné, aby vyhledávání probíhalo ve vlastních textech, pracuje tedy pouze s **indexovým souborem** (krátce **indexem**). Množství dokumentů, které je potřeba indexovat zase vyžaduje, aby tento proces probíhal nějakým způsobem automaticky. Primárními cíli vyhledávače jsou „úspěšnost“ (kvalita) a výkonnost (rychlost) [CROFT, 2012, s. 13].

Součástí každého vyhledávače je proto **robot** (crawler), který má za úkol procházet stránky a indexovat jejich obsah. Kvůli proměnlivosti webu probíhá crawling opakovaně, robot navíc prochází stránky podle nastavených priorit. Z těchto upřednostňovaných stránek extrahuje text, který je indexován (čímž vzniká zmíněný index), a hypertextové odkazy na další stránky, kam se poté také vydá.

Samotné **vyhledávání** probíhá tak, že rešeršní dotaz je konfrontován s indexem, jehož součástí jsou hypertextové odkazy. Následně jsou z něj extrahovány odkazy na stránky, které jsou relevantní (obsahují stejná slova jako rešeršní dotaz). Samotný rešeršní dotaz má obvykle podobu klíčových slov, lze použít i booleovské operátory nebo vyhledávat celé fráze.

Vyhledávače se snaží řadit výsledky podle **relevance**. Jak již bylo řečeno, v prostředí webu je třeba přihlížet i k jiným charakteristikám než je např. frekvence výskytu klíčových slov na dané stránce. Většina vyhledávačů si proto vytvořila nějakou metodu pro **hodnocení** (ranking) webových stránek. Příkladem může být Google PageRank, který se počítá pro každý web na základě počtu a důležitosti stránek, které na zkoumaný web odkazují [JANOVSKÝ, 2012].



## 2 Inteligentní vyhledávání

### 2.1 Vymezení pojmu

Definovat inteligentní vyhledávání není snadné. Podobně jako u vyhledávání informací obecně existuje mnoho definic a vycházejí přitom z různých přístupů.

Croft naznačuje, že **inteligentní vyhledávání** je vlastně jen jiný název pro „dobré“ vyhledávání informací. Tedy že připisovat inteligenci programu jako takovému není správné, protože inteligentní vyhledávání je každé vyhledávání, které je efektivní a výkonné [CROFT, 1987].

Naproti tomu existují přístupy, podle nichž je inteligentní vyhledávání realizováno počítačovým programem na základě instrukcí uživatele, ale bez jakýchkoli dalších interakcí s ním [MAES, 1994], nebo že inteligentní vyhledávání je využívání všech dostupných znalostí (o prohledávané bázi, jejích záznamech, obvyklých uživateli, zahrnutých dokumentech atd., stejně jako o konkrétním uživateli, jeho problému a požadavcích) počítačovým programem za účelem nalezení co nejlepšího „řešení“ uživatelova problému [BROOKS, 1987].

Brooks tedy říká, že aby vyhledávání probíhalo inteligentně, je potřeba, aby program pracoval se **znalostmi**. Využívání znalostí je ovšem typické pro člověka, nikoli pro počítačový program. To vedlo k nápadu zapojit do vyhledávání informací procesy, které si jinak spojujeme s lidskou inteligencí (porozumění jazyku, učení, řešení problémů, rozhodování vycházející ze zkušenosti apod.), tedy využít v rešeršních systémech inteligenci umělou [CHOWDHURY, 2004, s. 352-353].

Současným trendem reagujícím na informační explozi a zahlcení internetu redundantními informacemi je snaha zakomponovat hned několik technik umělé inteligence do jediného inteligentního agenta, kterého Gupta [2012] nazývá *intelligent information retriever*. Tento agent by „*lokalizoval, získával a zpracovával informace za účelem podpory rozhodovacího procesu.*“

## 2.2 Umělá inteligence

Pokud přistoupíme na to, že inteligentní vyhledávání nestojí pouze na inteligenci, kterou k tomuto procesu přispívá uživatel, je třeba představit v tuto chvíli inteligenci umělou.

Pojem umělá inteligence označuje celou vědní disciplínu, která se inspirovala v mnoha dalších oborech, jako je biologie, matematika nebo logika počítačového zpracování dat. Předmětem zkoumání této disciplíny jsou postupy (algoritmy), které simulují myšlenkové pochody a chování člověka [MAŘÍK, 1993, s. 15].

TDKIV prezentuje dva významy pojmu umělá inteligence [KUČEROVÁ, 2003]:

- 1. Mezioborová vědní disciplína na pomezí kognitivních věd, kybernetiky a počítačové vědy, která zkoumá a modeluje inteligenci s cílem vyvinout software a hardware, který bude při řešení úloh používat postupy považované za projev lidské inteligence. Základními oblastmi zkoumání jsou obecné řešení problémů, plánování, rozpoznávání, reprezentace znalostí, adaptace a strojové učení. V současné době jsou rozvinuty četné aplikační oblasti: expertní systémy, zpracování přirozeného jazyka, počítačové vidění, robotika, neuronové sítě ad.*
- 2. Vlastnost uměle vytvořených systémů, vykazující analogické rysy jako lidská inteligence.*

Umělá inteligence se dělí na psychologický a inženýrský směr. Psychologické zaměření se snaží na základě zkoumání lidské mysli a procesů v ní probíhajících napodobit její schopnosti strojově, třebaže po částech. V inženýrské větvi umělé inteligence je hlavním kritériem dosažený cíl, tedy vytvoření inteligentního stroje a to jakýmkoli způsobem [JORDAN, 1999, s. lxiii].

Důležitou součástí umělé inteligence byla a je robotika. Jedná se sice o samostatnou vědeckou a technickou disciplínu, ale je významnou aplikační oblastí umělé inteligence. Zabývá se úlohami plánování, v současnosti především úlohami plánování v reálném čase. Z počátku se jednalo o stacionární stroje, které řešily úzkou množinu úloh ve stabilním prostředí. Dnes je naopak zájem o mobilní, samostatné roboty, schopné řešit úlohy bez jakékoli externí pomoci [MAŘÍK, 1993, s. 26].

Dále se umělá inteligence aplikuje v počítačovém vidění (především metody rozpoznávání, metody reprezentace a využívání znalostí), při automatickém rozpoznávání a porozumění přirozenému jazyku, při zpracování dotykové informace, v lékařské diagnostice, v softwarovém a znalostním inženýrství a mnoha dalších oblastech.

Pokud jde o využití umělé inteligence ve vyhledávání, bude záležet na tom, ve které části vyhledávacího procesu se tato inteligence uplatní. Bude rozdíl, jestli to bude při konstrukci rešeršního dotazu, při jeho následném zpracování, při prezentaci výsledků nebo při komunikaci s uživatelem.

### **2.2.1 Dějiny umělé inteligence**

Historie umělé inteligence jako vědecké disciplíny není dlouhá, nicméně tato věda stihla od svého počátku projít nebývalým rozvojem. Počátky umělé inteligence se datují někdy kolem roku 1956, přesto již v roce 1950 přišel Alan Turing se svým testem a John von Neumann tvrdil, že je pravděpodobné, že počítače v dohledné době dosáhnou intelektuální úrovně člověka. K prvnímu užití termínu umělá inteligence došlo právě v roce 1956 na konferenci na Dartmouth College, New Hampshire, kterou zorganizoval John McCarthy spolu s Marvinem Minským, Nathanielem Rochesterem a Claudem Shannonem, a pozval na ni odborníky, zabývající se mentálními schopnostmi lidí i strojů. Ústřední myšlenkou bylo, že počítače mohou pracovat se symboly stejně jako s čísly, snažili se tedy dokázat, že ty principy lidské inteligence, které lze přesně popsat, je možné simulovat na stroji. Formulovali tak obsah nově vznikající vědní disciplíny, jejíž název navrhl právě John McCarthy [MAŘÍK, 1993, s. 16].

Ve stejném roce byl také představen první program s umělou inteligencí nazvaný *The Logic Theorist*. Vytvořili ho Allen Newell, J. C. Shaw a Herbert Simon a fungoval na počítači JOHNNIAC vytvořeném Johnem von Neumannem [LOGIC THEORIST, 2001]

Po těchto událostech se objevily prognózy, že v roce 1970 bude počítač velmistrem v šachu, že objeví nový matematický teorém, že porozumí přirozenému jazyku a bude sloužit jako překladatel, že bude komponovat hudbu na úrovni klasiků apod. Předpověď tehdy znamenala důležitý stimul v počátcích rozvoje umělé inteligence, bohužel se nakonec nevyplnila, což přineslo značné zklamání a tím i na nějaký čas ochabnutí zájmu o tuto vědní disciplínu [MAŘÍK, 1993, s. 19-20].

K výraznějšímu oživení došlo až v polovině 70. let, kdy byly vytvořeny úspěšné diagnostické expertní systémy MYCIN, který diagnostikoval infekční onemocnění krve, a PROSPECTOR, který odhaloval rudná ložiska. PROSPECTOR si velmi záhy po svém představení získal popularitu, protože odhalil ložisko molybdenových rud ve státě Washington v hodnotě 100 milionů dolarů [MAŘÍK, 1993, s. 21]. Umělá inteligence tedy začala pracovat se znalostmi, což v 80. a 90. letech následuje snaha úlohy přehledně modelovat. Ve stejné době se také rozvíjela myšlenka strojového učení, získávání znalostí z dat a následně i jejich tzv. dobývání.

Současným trendem ve využívání znalostí je tvorba ontologií (v podstatě formalizovaných terminologických map pro konkrétní oblasti, které usnadňují komunikaci a šíření znalostí dané problematiky) a sémantického webu, který se snaží začlenit do webových stránek sémantický obsah a prezentovat tak na webu strukturovaná data s jasným významem.

### **2.2.2 Artificial general intelligence**

Pojem *artificial general intelligence* (AGI), jinak také *strong-AI* nebo *full-AI* se do češtiny někdy překládá jako umělé bytí nebo umělý život. Jedná se o dlouhodobý cíl výzkumu v oblasti UI, jenž se snaží vytvořit stroje, které dokážou myslet, tedy takovou umělou inteligenci, která by byla v podstatě k nerozeznání od inteligence lidské [ARTIFICIAL INTELLIGENCE (AI), 2014]. Je to, alespoň zatím, vlastně jen vyslovená hypotéza. Dosud se ani neví, je-li něco takového vůbec možné, natož aby se tzv. umělé bytí někomu skutečně podařilo vytvořit.

Snaha vyrobit takový stroj je pochopitelná vzhledem k tomu, že zatím všechny aplikace prvků umělé inteligence jsou v podstatě jen dílčí. Významným rysem inteligence je přece také všestrannost a přizpůsobivost a těch se na jednotlivostech dosáhnout nedá.

Z hlediska IIR je *strong-AI* pochopitelně zajímavá – jestli by dokázala fungovat jako plnohodnotný člověk, rozhodovat se, učit se, vyvozovat závěry, a přitom fungovat bez omezení v podobě rychlosti, únavy a podobně, byla by dokonalým nástrojem pro vyhledávání (a samozřejmě nejen pro něj). Je ovšem otázka, jestli by to bylo vlastně možné, i kdyby *strong-AI* opravdu existovala. Dá se předpokládat, že pokud by vlastně byla „bytosť“, uvědomovala si sama sebe a měla

schopnost cítit, měla by také vlastní vůli. Možná by se dalo tedy říct, že jako taková je to skutečná živá bytost, byť uměle stvořená, nikoli pouhý stroj.

*AGI* představuje tedy i filozofický, respektive etický problém, a její existence by pro lidstvo, které se dosud (pokud víme) nesetkalo s jinou podobně inteligentní formou života, představovala vlastně velký problém. Jakým způsobem se k ní postavit, když se jedná o formu života, která je navíc lidským výtvořem? Člověk sice má schopnost se učit, ale často se bohužel učí až ze svých chyb. A v případě *AGI* zatím neměl možnost chybovat.

### **2.3 Inteligentní prvky ve vyhledávání informací**

Pokud budeme pracovat s nejobecnější definicí inteligentního vyhledávání, kterou formuloval Croft [1987], tedy že inteligentní je každé vyhledávání, které je „dobré“, lze určit některé prvky, které ho „dobrým“ činí.

Když ideu inteligence ořežeme na úplný základ, jedná se vlastně o schopnost naplánovat nějaký úkon tak, aby pak jeho provedení zabralo co nejméně času při nejvyšší možné efektivitě. Takovéto plánování probíhá na základě mnoha faktorů od předchozích zkušeností té dané osoby s problémy podobné povahy, přes aktuální podmínky a dostupné zdroje, až po účel, za jakým osoba danou informaci shání. Otázkou zůstává, zda je vůbec možné přimět podobně pracovat počítačový program. Inteligentní vyhledávání pochopitelně počítá s tím, že to alespoň částečně možné je – inteligence daného vyhledávání nemusí spočívat přímo v jádru programu nebo ve vyhledávacím algoritmu, ale může se odrážet v uživatelském rozhraní, ve způsobu komunikace s uživatelem, v systému nápověd.

Pokud člověk shání například nějakou konkrétní informaci týkající se dějin hudby, zcela automaticky vyloučí ze své cílové skupiny zdrojů přírodovědné encyklopedie. Naopak „hloupý“ program by v takovém případě otrocky procházel všechna dostupná data. Není náhoda, že jednou z významných typologií databází je příslušnost k oboru.

Na jednu stranu se může stát, že neelegantní přístup programu k problému přinese zcela nečekané, až objevné výsledky, kdy dá do souvislosti něco, co by člověka v takové situaci třeba vůbec nenapadlo. Faktem ale zůstává, že takové

výsledky jsou dílem náhody. Ta sice, jak můžeme vidět na četných příkladech z historie lidského poznání, měla podíl na mnoha vědeckých objevech, nicméně neefektivně vyhledávací proces jen proto, že v jednom z miliardy případů přinese něco nového, je krajně nevýhodné.

### **2.3.1 HCI a uživatelské rozhraní**

Obor zabývající se problematikou komunikace člověk-počítač se nazývá HCI, což je zkratka anglického termínu *human computer interaction*. Souvislost s informační vědou potvrzuje zařazení termínu do TDKIV, která ho definuje jako *„druh komunikace, při níž dochází k přenosu informací mezi člověkem a počítačem, která spočívá v interakci programátora, operátora či uživatele s počítačem na základě přesně stanovených pravidel. Vstupní informace jsou počítači předávány např. pomocí klávesnice, hlasového vstupu apod. Výstupní informace předává počítač člověku pomocí monitoru, tiskárny, hlasového výstupu apod.“* [JONÁK, 2003].

Je logické, že pro takový směr řešení problému je nutné, aby HCI využívalo poznatky ze zkoumání obou stran zmíněné interakce, tedy ze strany uživatele i počítače. HCI je tak typicky průnikovým oborem. Dává dohromady poznatky z psychologie kognitivních věd, sociologie a dalších oborů, a zároveň využívá informační technologie a snaží se najít optimální způsob, kterým by s dostupnými technologiemi byla komunikace co nejjednodušší a přitom pro uživatele intuitivní a efektivní. Koncový uživatel stojí ve středu HCI, je zde klíčovým prvkem.

Důležitým předmětem zkoumání HCI je uživatelské rozhraní, jehož vývojem a designem se (mimo jiné) HCI zabývá. TDKIV je charakterizuje jako *„rozhraní mezi uživatelem a počítačovým programem, které mj. zahrnuje možnosti a postup ovládání programu, chybová hlášení programu, formu a obsahu nápovědy apod.“* [POTÁČEK, 2003]. Uživatelské rozhraní se skládá ze dvou hlavních částí [PAPÍK, 2001] :

- *Fyzická část (např.: vstupní/výstupní zařízení, nástroje selekce a zpětné vazby)*
- *Konceptuální část (dotazovací, resp. příkazové jazyky, menu, prostředky přímé manipulace, systémy pracující s ikonami, formulářové rozhraní, prvky inteligentního rozhraní)*

Někdy se uvádí také část perceptuální, ve které jde například o vnímání uspořádání objektů na obrazovce nebo použití barev [SOUČKOVÁ, 2003].

Důležitými principy v dobrém uživatelském rozhraní jsou informativní zpětná vazba, snadná reverzibilita kroků, uživatelské podpora, poskytování různých druhů rozhraní podle míry zkušeností uživatele aj.

Pokud jde o zpětnou vazbu uživateli, měla by se týkat jednak vztahu mezi uživatelským rešeršním dotazem a vyhledanými dokumenty, a jednak vztahů mezi dokumenty samotnými. Rozhraní pro různě informačně gramotné uživatele je zase důležité pro zachování všech možností vyhledávání – s jednoduchými rozhraními se uživatelé snáze naučí zacházet, ovšem za cenu menší flexibility a výkonu. Je proto výhodné mít i rozhraní pro zkušené uživatele, kteří dokáží plně využít i jeho pokročilé funkce a kteří přitom nemají pocit, že jsou nuceni učit se složité operace a porozumět podrobně fungování daného rešeršního systému (což nenároční uživatelé rozhodně dělat nechtějí).

Inteligentní rozhraní pro vyhledávání informací nicméně nemá přesně stanoveny prvky, které by mělo obsahovat, ani předem danou funkcionalitu; neexistuje přesný teoretický model. Základem je, že se snaží propojit poznatky z HCI, vyhledávání informací a umělé inteligence [MIZZARO, 1996, s. 2]. Za inteligentní prvky v uživatelském rozhraní můžeme považovat ty, které si spojujeme spíše s komunikací s člověkem než s počítačem, jako je vnímání obrazu, zvuku, opravování chyb, navrhování alternativních řešení apod. Dobře navržené a efektivní uživatelské rozhraní pak vlastně vypadá tak, že pokud s ním uživatel správně zachází, rozhraní v podstatě „zmizí“ a umožní mu koncentrovat se výhradně na práci.

### **2.3.2 „Inteligentní chování“ uživatelského rozhraní**

Jestliže se rešeršní systém chová v komunikaci s uživatelem podobně, jak by se v takové situaci choval člověk, tedy nepřijímá slepě podněty jen jako příkazy a úkoly, ale reaguje na ně a nabízí uživateli další možnosti, dá se říci, že se chová „inteligentně“. Nezanedbatelným prvkem inteligence rešeršního systému je jistě schopnost prezentovat uživateli veškerou svou funkcionalitu, možnosti a nástroje tak, aby jim uživatel pokud možno bez větší námahy dokázal porozumět a plně tak využít škálu služeb, které daný rešeršní systém nabízí. Běžným problémem totiž bývá, že systém sice má mnoho možností, jak modifikovat rešeršní dotaz, jak filtrovat

výsledky, nebo ve kterých oblastech vnitřní báze vyhledávat, nicméně uživatel o nich nemá ani ponětí.

Důležitým faktorem je relevantní zpětná vazba, která nemusí sloužit jen k bližší specifikaci rešeršního dotazu, ale i jako ukazatel zkušenosti uživatele se systémem. Uživatel ve zpětné vazbě označí jako relevantní dokumenty jen ty, které jsou v češtině nebo němčině. Systém na základě toho může uživateli představit možnost, jak výsledky vyhledávání limitovat jazykově rovnou při konstrukci dotazu, případně nabídnout použití nástroje pro instantní překlad stránek v nepreferovaných jazycích, pokud ho má (toto funguje obvykle na internetu při vyhledávání v povrchovém webu, kdy má vyhledávač přístup k plnému textu stránek).

### **2.3.3 Počítačové vidění**

Počítačové vidění je schopnost počítače vnímat, zpracovat a interpretovat obrazové vjemy. Jedním z jejích cílů je naučit stroj ze vstupních dvourozměrných obrazů extrahovat informaci o trojrozměrném prostoru, který zachycuje [MACHINE VISION, 1999, s. 501]

V této oblasti byly již vyvinuty softwary, např. nástroj *Emotient*, které dokážou rozpoznávat lidské emoce na základě výrazu tváře a to jak z videí nebo obrázků, tak v reálném čase z živého vstupu pomocí kamery [EMOTIENT, 2014]. Pro komunikaci člověk-stroj při vyhledávání informací je to minimálně zajímavé – počítač může odhalit reakci uživatele na prezentované výsledky, například překvapení, rozčarování nebo naopak radost, a na jejím základě vlastně sám od sebe získat zpětnou vazbu. Samozřejmě problém, jako se všemi simulacemi lidského chování strojem, je v tom, že nic není jednoznačné. Lidé jsou individuality, a přestože rámeček je možná stejný, každý projevuje své emoce s nějakou alespoň drobnou odlišností. Člověk je přesto schopen snadno poznat, jak se druhá osoba cítí díky zkušenosti a v neposlední řadě instinktu, což je fenomén, který se do počítače implementovat zatím ani nedá, protože mu sami ještě úplně nerozumíme.

Počítačové vidění se podobně jako všechna odvětví strojové percepce inspirovalo v neurovědě. Lidský mozek vyhodnocuje aktuální náladu druhého člověka mimo jiné také z výrazu tváře. Program pochopitelně funguje trochu jinak, obsahuje sadu přesně definovaných parametrů a jejich kombinací spolu s určením míst, kde je má sledovat. Podle nich se učí rozpoznávat jednotlivé emoce. Patří sem



například vrásky na čele, kolem očí a úst, pohyby očí a očních víček, pohyb úst a rtů, zabarvení tváří atd. Jiné charakteristiky (řeč těla, hlasové projevy, okolí uživatele aj.) na rozdíl od člověka zohlednit neumí.

Hlaváč [1997, s. 178-179] uvádí, že pro interpretaci vizuálních dat je potřeba o objektu předem něco vědět, disponovat nějakou apriorní znalostí. Tu ovšem počítače postrádají. Pokud bychom například dali zmíněnému programu *Emotient* k rozpoznání emocí obrázků nějakého primáta s hrubými rysy dostatečně podobnými těm lidským (gorila, šimpanz), nepoznal by to, analýzu by provedl a vrátil by nějaký zkreslený výsledek.

V současné době není běžné, aby uživatelské rozhraní rešeršních systémů integrovalo počítačové vidění. Někdy dokáže zpracovávat vizuální vstupy, především obrázky, ale nejedná se o počítačové vidění. Systém sice dokáže z obrázku analýzou zjistit mnohé z jeho charakteristik, nicméně neumí je interpretovat. Dovede na jejich základě například vyhledat podobné obrázky nebo je zjištěná data porovnat s charakteristikami v databázi a určit, o jaký obrázek se jedná (za předpokladu, že se v databázi nachází, např. u výtvarných uměleckých děl), ale jestli je na obrázku loď nebo strom, to rozlišit nedokáže.

### **2.3.4 Počítačový sluch**

Podobně jako počítačové vidění se soustřeďuje na porozumění obrazovým informacím spíše než na jejich zpracování, cílem „počítačového sluchu“ je porozumění zvukovým informacím. Člověk dokáže interpretovat informaci i v případě ztížených komunikačních podmínek – opět díky apriorním znalostem, a to znalostem obecným o světě kolem nás, ale především o tématu konverzace samotném [PSUTKA, 2007, s. 285]. Aby byl něčeho takového schopen i počítač, musí nutně mít přístup do nějakého zdroje znalostí souvisejícího s daným tématem. Chceme-li alespoň podobné úroveň porozumění dosáhnout i při strojovém zpracování zvukových vjemů, je nutné přesně analyzovat zvolenou úlohu a modelovat potřebné zdroje znalostí [MINKER, 2002].

Pokud má být hlasová komunikace obousměrná, počítač potřebuje nejen nástroj na rozpoznávání řeči ale také nástroj na její syntézu. Zpracování přirozeného jazyka je obecně velmi komplikovaná problematika, a tak aplikace hlasové komunikace se stále soustřeďuje na konkrétní aplikační oblast (nákup po telefonu,

rezervace jízdenky), nikoli na přirozenou a neomezenou konverzaci na libovolné téma [PSUTKA, 2007, s. 284].

Jako jednoduchý nástroj k hlasovému ovládní počítače může sloužit např. software *e-Speaking*. Zahrnuje přes 100 předdefinovaných povelů, díky nimž je možné ovládat myš, otevřít e-mail, pouštět programy nebo dokonce zaznamenávat diktát. V případě zakoupení licence umožňuje i přidávání vlastních povelů [E-SPEAKING, 2013]. Další typickou aplikací této technologie je hlasové vytáčení v mobilních telefonech. Ve vyhledávání informací používá počítačový sluch například internetový vyhledávač *Google*, a to ke konstrukci rešeršního dotazu.

### **2.3.5 Počítačový hmat**

Počítačový hmat je oblast zkoumání schopností strojů vnímat a interpretovat taktilní informace. Hmat, jakožto jeden z lidských smyslů, nám umožňuje získávat informace o svém okolí, aniž bychom si to třeba uvědomovali, a jako takový může být podceňován. Jeho význam je ale značný, zvláště pokud je nejvýznamnější z našich smyslů, zrak, nějakým způsobem znehodnocen.

Hmatové informace získáváme dvěma různými způsoby: osaháváním povrchu předmětu a uchopením celého předmětu. Obou dvou způsobů lze využít také u strojů. Kromě hmatových receptorů se v lidské kůži nacházejí také termoreceptory, které prostřednictvím změn teploty kůže v místě dotyku reagují na teplotu předmětu. Termoreceptory ale nejsou v aplikaci u strojů zase tak významné, simulovat je při technické realizaci se ovšem někdy také snažíme – v případech, kdy zamýšleným příjemcem strojového hmatu není robot, ale člověk, který nějak o vlastní hmatové receptory přišel.

Hlavní oblast aplikace počítačového dotyku je v robotice a oborech, kde se robotika využívá, např. ve strojírenství. Taktilní snímače mohou sloužit k rozpoznání tvaru a rozměru předmětu, k vyhodnocení síly stisku, kterým stroj může předmět uchopit, k monitorování okolí a následného varování v případě přiblížení neočekávaného objektu (tedy jako bezpečnostní opatření), nebo k orientaci stroje v prostoru [VOLF, 2007, s. 349].

Počítačový hmat asi nebude středem zájmu integrace inteligentních prvků do rešeršních systémů, nicméně dovedu si představit jeho využití např. při prohledávání databází různých 3D objektů nebo stavebních materiálů, kdy by systém byl schopen

dotykem zjistit např. hrubost, tuhost, prokluz atd., určit tak druh a vlastnosti představovaného materiálu.

Co se ovšem již v docela hojně míře používá, jsou dotykové displeje, které neslouží k interpretaci hmatové informace jako takové, nýbrž jako snímače povelů. Dotyk, tentokrát ovšem ze strany uživatele, znamená vždy konkrétní instrukci v závislosti na tom, jak a kde v rozhraní byl proveden. Implementace je pochopitelně přímo závislá na hardwarovém vybavení a v podstatě nijak nesouvisí s vlastnostmi rešeršního systému jako takového.

### **2.3.6 Strojové učení**

S umělou inteligencí úzce souvisí pojem strojového učení, který se právě odráží od biologické inteligence: „*Důležitou vlastností živých organismů je schopnost přizpůsobovat se měnícím se podmínkám (adaptovat se), eventuálně učit se na základě vlastních zkušeností. Je přirozené, že vybavit touto vlastností i systémy technické je jedním z cílů umělé inteligence.*“ [MAŘÍK, 1988, s. 2].

Když to značně zjednodušíme, dá se říct, že učení u stroje probíhá vždy ve chvíli, kdy nějakým způsobem modifikuje buď sám sebe, své nastavitelné parametry nebo data v sobě obsažená, a činí tak na základě vstupů v reakci na přijaté informace.

Langley a Simon identifikují pět hlavních oblastí výzkumu strojového učení, přičemž o čtyřech z nich se už v souvislosti s vyhledáváním informací mluvilo v různých studiích. Jsou to: rozhodovací stromy a indukce, učení založené na instancích, neuronové sítě, genetické algoritmy a analytické učení [LANGLEY, 1995].

Metody učení jsou buď induktivní nebo deduktivní. Induktivní metoda znamená vyvození všeobecně platného závěru z pozorování množiny jevů (syntetický přístup k učení), zatímco deduktivní metoda vychází z jediného jevu, respektive z jeho analýzy (analytický přístup k učení) [KUBÁT, 1993].

Učení, např. u neuronových sítí, může probíhat za účasti učitele, tj. nějakého vnějšího činitele, který zná správnou reakci na každý stimul, nebo – bez znalosti učicí množiny – i bez něj. Ideální stav je, když po absolvování takového tréninku a nacházení souvislostí v trénovacích datech, dokáže daný systém řešit úlohy i jen lehce

podobné, s úplně jinými vstupními daty. Tedy podobně jako živé organismy – adaptovat se na změnu.

Berka [2003] adaptivitu představuje „jako inkrementální učení (průběžné doučování systému na základě nových dat), učení a zapomínání (schopnost systému rozpoznat změnu konceptu a této změně se přizpůsobit), integraci znalostí (schopnost systémů kombinovat více zdrojů znalostí), meta-učení (schopnost systémů kombinovat více modelů znalostí v průběhu rozhodování), revizi znalostí (schopnost aktualizovat používané znalosti), nebo využívání analogií.“ Adaptivita je významným znakem inteligentního chování a její dosažení je dlouhodobým cílem umělé inteligence.

### **Strojové učení ve vyhledávání informací**

První čtyři výše jmenované oblasti strojového učení, tedy rozhodovací stromy a indukce, učení založené na instancích, neuronové sítě a genetické algoritmy, se učí z informací s jednoduchou strukturou – z příkladů konceptu, který se mají naučit, často znázorněného symbolickými nebo číselnými schématy. Pro generování struktury, která reprezentuje vztahy zahrnuté v obsažených datech, se pak používají heuristické metody. Tyto oblasti strojového učení se využívají právě v inteligentních rešeršních systémech, kde jsou dokumenty reprezentovány pomocí jednoduchých měřítek, jakým je např. frekvence slov.

Poslední, pátá oblast strojového učení, analytické učení, se obvykle používá k učení důkazů, respektive vysvětlení pro příkladové situace díky nějaké znalosti, která je v systému implementována. Tím, že systém pak seskupí poznatky z několika takových ukázkových případů, může být schopen redukovat čas i prostředky pro hledání řešení podobné situace v budoucnu. Bohužel potřeba vložených znalostí a složitá struktura daná nutností shromažďovat někde zmíněná vysvětlení, dělají z analytického učení metodu nevhodnou pro vyhledávání ve velkém měřítku.

Ve vyhledávání informací se uplatňuje učení [HOFMANN, 2006]:

- bez dohledu (proces interpretace), kdy systém sám pracuje s daty v sobě obsaženými a objevuje skryté zákonitosti, generuje sémantické reprezentace, využívá postupy prediktivního modelování, vytváří statistiky atd. (patří sem například data-minig)

- pod dohledem (proces generalizace), kdy systém generalizuje procesy z jednotlivých příkladů, na jejich základě data klasifikuje a rozpoznává a napodobuje tak činnost lidského experta

Pro vyhledávání informací může být tato schopnost velkým přínosem. Představa, že čím více dotazů systém zpracuje, tím se stává efektivnějším, zvyšuje se jeho výkon a rychlost, je skutečně lákavá. Ideální by bylo, pokud by se systém učil kontinuálně, nejen v přípravném procesu s trénovacími daty nebo v konkrétních případech, kdy získaná „zkušenost“ nezůstává v systému trvale (relevantní zpětná vazba).

Použití učicích algoritmů k automatizaci některých procesů z vyhledávání informací jako je např. klasifikace dokumentů může nejen značně urychlit zpracování dat a ušetřit lidskou práci, ale také snáze kompenzovat určitou nekonzistentnost a chybovost, která je pro člověka typická. Strojové učení nalézá uplatnění také ve vyhledávání netextových informací, kdy se používá k rozpoznávání vzorů např. při vyhledávání hudebních informací.

### **2.3.7 Zpracování přirozeného jazyka**

KTD definuje přirozený jazyk jako: „*Jazyk, který se vyvíjí, jehož pravidla vyplývají z úzu (nejsou definována formálně) a jenž je užíván konkrétním jazykovým společenstvím (např. národní jazyky)*“ [BALÍKOVÁ, 2003]. Je to otevřený systém, jehož univerzálnost nám umožňuje vyjádřit pomocí něj prakticky cokoliv, co se rozhodneme někomu sdělit. V případě potřeby si vhodné prostředky můžeme v jazyce vytvořit sami způsobem, který nám daný jazyk umožňuje [HAJIČOVÁ, 1997, s. 244]

Už z názvu pojmu vyplývá, že pro člověka je přirozené používat tento jazyk a je tedy nasnadě, aby inteligentní systém byl schopen tomuto jazyku porozumět, analyzovat data jím formulovaná a podobně. Zpracování a porozumění přirozenému jazyku je tedy důležitou součástí umělé inteligence a pro inteligentní vyhledávání má značný význam.

Jevy, které plynou z historického vývoje jazyka a jsou typické pro všechny jeho varianty (národní jazyky), jsou pro umělou inteligenci skutečnými úskalími. Například homonymie a polysémie, tedy existence totožného výrazu pro označení různých věcí, kdy lidská inteligence je bez jakékoli námahy schopna přiřadit danému

pojmu ten správný význam dle kontextu a situace, jsou pro umělou inteligenci velkým problémem.

Rozlišení významů slov (anglicky *word sense disambiguation*, WSD) je jedna z klíčových akcí, kterou musí systém pro zpracování přirozeného jazyka udělat. Systém víceznačná slova roztřídí do skupin, které představují vždy jen jeden význam slova. Využívá k tomu například slovníky, kde jsou jednotlivé významové možnosti popsány i včetně obvyklých doprovodných indikátorů, které pomohou význam správně přiřadit [LAPATA, 2007]. Například v automatickém překladu je rozlišení významů, celkem logicky, nepostradatelné.

### **Zpracování přirozeného jazyka do strojem čitelné podoby**

Samotní porozumění přirozenému jazyku je teprve prvním krokem v jeho strojovém zpracování. Důležitou součástí tohoto procesu je totiž také vytvoření výstupu v nějakém strojově čitelném jazyce, tedy v nějaké určité struktuře. Právě tento výstup pak slouží jako základ pro sémantickou interpretaci nebo pro další procedury (strojový překlad aj.) [HAJIČOVÁ, 1997, 245 s.]. Aby toho stroj docílil, provádí analýzu vstupních dat. K nejčastějším typům analýzy patří morfologická, syntaktická a sémantická analýza [HABROVANSKÁ, 2010]:

- Morfologická analýza – zabývá se jednotlivými slovy a určuje slovní druhy, tvary a další gramatické kategorie.
- Syntaktická analýza – zkoumá větné celky, u nichž provádí rozbor větných členů a závislostních vztahů mezi nimi
- Sémantická analýza – snaží se zachytit význam vstupního textu. V podstatě značkuje jednotlivé výrazy v přirozeném jazyce, přičemž tyto značky specifikují význam těchto výrazů.

## **2.4 Terminologická analýza**

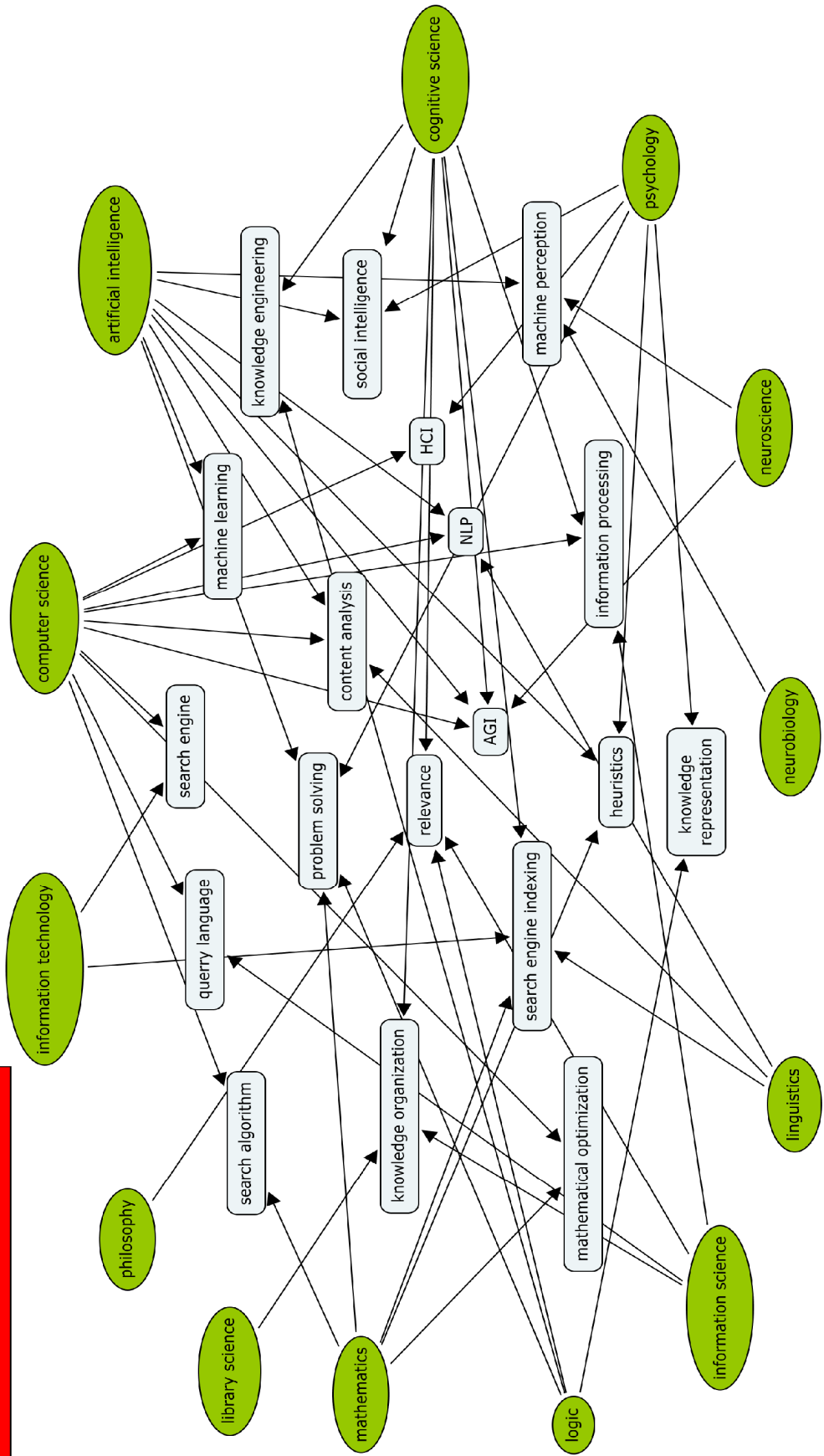
V anglické verzi otevřené webové encyklopedie *Wikipedia* [2001] jsem na základě odborné literatury citované v této kapitole provedla průzkum termínů souvisejících s inteligentním vyhledáváním. Samostatné heslo „inteligentní vyhledávání“ („intelligent information retrieval“) není ve *Wikipedii* zahrnuto.

*Wikipedii* jsem vybrala proto, že je masově využívána, velmi obsáhlá a díky otevřenosti stále se rozšiřující encyklopedie, která má (ve své anglické verzi) i přes zmíněnou otevřenost svá hesla kvalitně zpracovaná. Právě díky té otevřenosti zpracovává relativně rychle nová témata, což je v oblasti inteligentního vyhledávání důležité.

Na základě vnitřního začlenění termínů do oborů v této encyklopedii jsem vytvořila vizualizaci svého průzkumu. Oborové začlenění termínů se tedy opírá o vlastní vnitřní strukturu *Wikipedie*. Pro účely průzkumu termínů jsem vycházela z nejjednoduššího výkladu pojmu inteligentní vyhledávání, totiž že se jedná o vyhledávání informací (information retrieval) s využitím umělé inteligence (artificial intelligence). Skrze analýzu těchto dvou hlavních hesel, jsem vytvořila jakousi mapu pojmů, které s inteligentním vyhledáváním souvisejí, a na jejich základě sestavila zmíněné grafické znázornění podle oborů, do kterých inteligentní vyhledávání zasahuje (obr. č. 1).

Pojmy jsou uvedené v originálním znění, protože pro některé z nich neexistuje zažitý český ekvivalent. Rozhodla jsem se tedy ponechat je všechny v angličtině. Uspořádání termínů je vytvořeno tak, aby mapa byla přehledná; jejich rozmístění a vzdálenost od názvů jednotlivých oborů nevyjadřují, jaký k nim mají vztah.

**Intelligent information retrieval**





**Obr. č. 1. Schéma oborového začlenění inteligentního vyhledávání podle  
*Wikipedie***

### **3 Použití inteligentního vyhledávání**

#### **3.1 Vyhledávání obrazových informací**

Díky rozvoji internetu a možnostem, které nám současné informační technologie přinášejí, se rozrostlo množství formátů informací, které je možné pořádat, indexovat a následně v nich vyhledávat. Vyhledávání v netextových dokumentech jako jsou obrázky, hudba nebo multimédia je oblastí, v nichž má možnost se inteligentní vyhledávání nejvíce uplatnit.

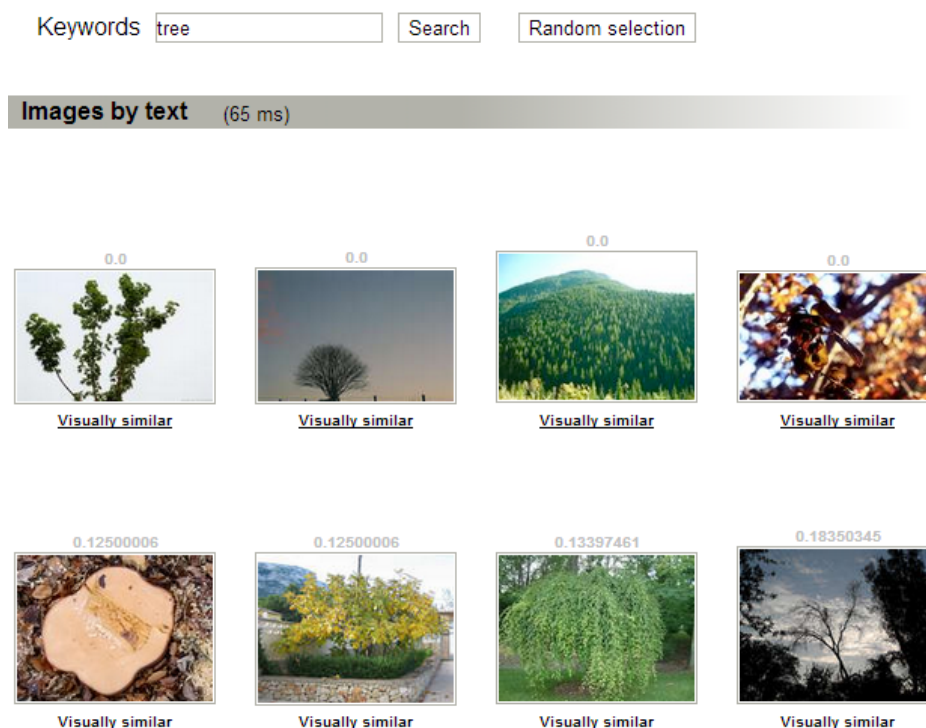
Nejmarkantněji je tento trend vidět v prostředí internetu – dnes je možné prakticky ihned při pořízení fotografie umístit ji na internet a sdílet, integrace fotoaparátu do mobilního telefonu je naprosto běžnou záležitostí, připojení k internetu přes mobilní síť rovněž. Je nasnadě, že se záhy objevily metody, jak v obrázcích vyhledávat.

Prvními byly metody využívající slovního popisu obrázků, nebo vyhledávání v textech, které obrázky doprovázely. Také se uplatnilo tzv. tagování, což je vytváření metadat samotnými uživateli – vzhledem k povaze takového popisu obrázků je rozšířené spíše na volném internetu než v uzavřených databázích. Využívá se na serverech pro sdílení fotografií jakým je například služba *Picasa* od společnosti Google.

Novějším způsobem jak vyhledávat obrázky je na základě analýzy obsahu podle parametrů jako jsou barva, tvar, zaplnění plochy apod. Jedná se tedy o technický obsah, nikoli obsah významový.

Vyhledávač obrázků od firmy Google dokáže vyhledávat jak podle slovního popisu obrázků, tak přímo podle obrázku – služba umožňuje nahrát vlastní soubor nebo odkázat na již zveřejněný online obrázek, který lze pak také doplnit ještě slovním popisem. V případě použití vzorového obrázku pro vyhledávání se ve výsledcích zobrazí obrázky podobné původnímu, odkazy na webové stránky, které původní obrázek obsahují, a další velikosti původního obrázku [GOOGLE, 2001b].

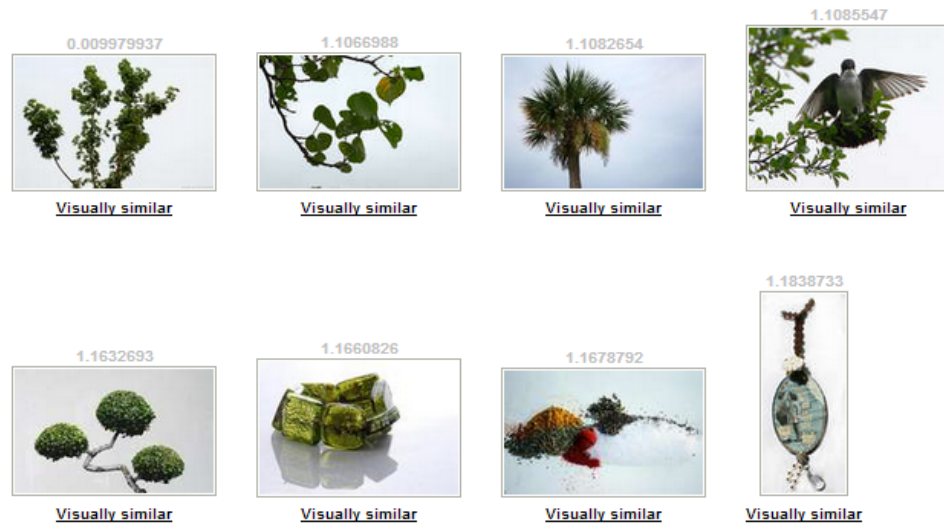
Podobnostní vyhledávání obrázků je směr, který se v současné době dále rozvíjí. Fakulta informatiky Masarykovy univerzity v Brně se například podílí na projektu *MUFIN – Multi-feature Indexing Network* (<http://mufin.fi.muni.cz/>), který obsahuje i demo pro podobnostní vyhledávání obrázků. Dlouhodobým cílem projektu je vyvinout univerzální technologické řešení problému vyhledávání v různorodých a mohutných databázích. Vyhledávač funguje nad databází Flickr (Flickr je služba pro nahrávání a sdílení fotografií), která používá k indexaci fotografií uživatelské tagy. Po zadání klíčového slova, tagu, do dialogového okna systém vrátí výsledky, které tento tag (tagy) obsahují (obr. č. 2). Následně je možné vybrat si z výsledků preferovaný obrázek a volbou „visually similar“ si nechat zobrazit fotografie, které jsou mu vizuálně podobné bez ohledu na tagy (obr. č. 3).



**Obr. č. 2: Rozhraní vyhledávače *MUFIN* a část výsledků pro dotaz „tree“**

Z hlediska inteligentního vyhledávání je nejzajímavější třetí možnost, kterou Google obrázky nabízí, a tou je využití hlasového vstupu. Stačí mít v počítači integrovaný mikrofón, případně ho připojit externě. Po zvolení této možnosti vyhledávání se objeví velká blikající ikona mikrofónu a výzva na uživatele, aby zadal povel. Rozpoznávač řeči reaguje jen na anglické povely a mě tedy nepřesvědčil, že by se mohl stát mou preferovanou formou formulace rešeršního dotazu. Mohlo to

samozřejmě být mým hardwarem, nicméně konstrukce dotazu se opakovaně nedařila, mikrofon registroval přílišné množství šumu a dekodér nedokázal rozlišit hlasový projev. Když už se to nakonec povedlo, stejně v drtivé většině případů vyhledávač rozuměl špatně (místo „dog“ hledal „do“ apod.).



**Obr. č. 3: Výsledky podobnostního vyhledávání ve službě *MUFIN* po výběru první fotografie na obr. č. 1 jako vzorové**

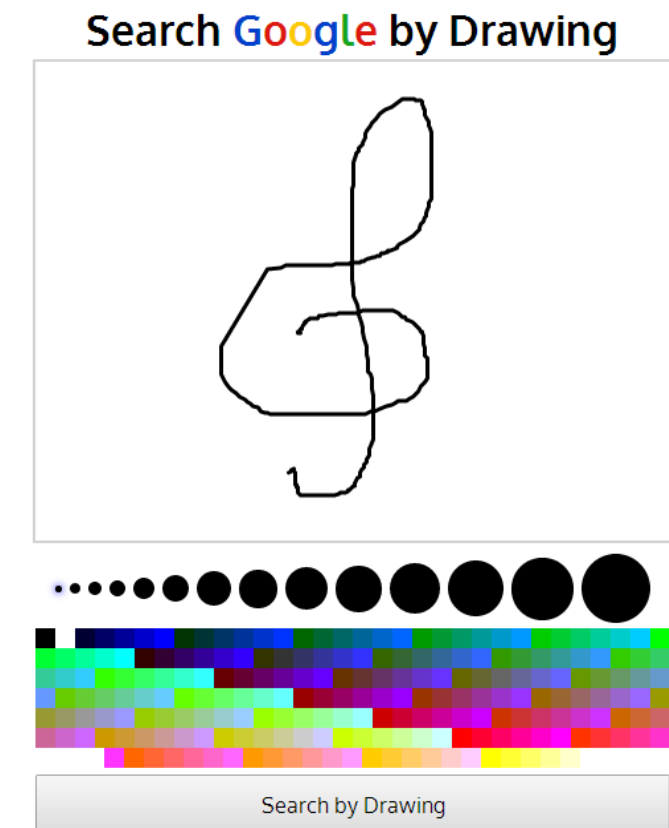
Google jako otevřený vyhledávač zpracovává obrovské množství obrázků, jeho index je ohromný. Obrázkové vyhledávání se ale uplatňuje i v uzavřených systémech jako jsou obrazové digitální knihovny a digitální sbírky například uměleckých muzeí.

### **3.1.1 Vyhledávání obrázků kreslením**

Na internetu funguje také neoficiální služba Googlu dříve nazývaná *Sketch-Based Image Search*, nyní prostě pojmenovaná *Search Google by Drawing* (<http://search-by-drawing.fullstackoptimization.com/>), která by se dala popsat jako vyhledávání obrázků kreslením. Jedná se ale o nezávislý projekt jednoho člověka, Franze Enzenhofera, který pouze funguje nad googleovským indexem obrázků; Google tuto službu nezařičuje.

V tomto druhu vyhledávání spočívá formulace rešeršního dotazu v nakreslení vlastního obrázku v nástroji, který je podobný klasickému programu *Malování* od firmy Microsoft, jen s méně funkcemi – je možné vybírat pouze z barvy a tloušťky

čáry (obr. č. 4). Nástroj následně porovnává charakteristiky, jako jsou zakřivení čar, tvary, barvy, síla čáry atd. se svým indexem obrázků.



**Obr. č. 4: Pracovní rozhraní nástroje *Search Google by Drawing***

Díky práci s googlovským indexem se po odeslání dotazu zobrazí známé rozhraní s výsledky a uživatel má možnost doplnit svůj rešeršní dotaz o slovní popis. Jak je patrné při porovnání nejbližších výsledků a rešeršního dotazu na obr. č. 5, míra shody není valná.

Problémy, které se s tímto způsobem tvorby dotazu mohou vyskytnout, jsou pochopitelně buď technického rázu (nástroj zdaleka nefunguje ve všech prohlížečích a na všech zařízeních, což je ovšem u neoficiální služby pochopitelné), nebo se týkají nástroje samotného – například v situaci, kdy uživatel nechce, aby barva obrázku hrála roli. Takto dotaz limitovat nelze. Další velkou nevýhodou je pochopitelně doba konstrukce rešeršního dotazu. Snad s přívětivějšími prostředky než jsou myš nebo touchpad by se to dalo zvládnout v rozumném čase. Pro většinu dotazů je ale tento způsob vyhledávání spíše nevýhodný.



Velikost obrázku:  
480 × 360

Obrázek nebyl nalezen v jiných velikostech.

Tip: Zkuste do vyhledávacího pole zadat popisné slovo.

Vizuálně podobné obrázky

Nahlásit obrázky



**Obr. č. 5: Rešeršní dotaz a část výsledků vyhledávání v nástroji  
*Search Google by Drawing***

I tak mě napadají příklady, kdy může být docela užitečný – představme si, že máme nějaký symbol, logo nebo něco podobného a nevíme, co znamená. Něco takového je obtížné popsat slovy, když tato vlastně nevyjadřují vlastní význam symbolu. Například takový symbol pro jin-jang nebo různá firemní loga se popisují slovy obtížně, ale nakreslit je není takový problém.

Nástroj může být šikovný také pro vyhledávání strukturních chemických vzorců, ale to spíš jen pro náhodné zájemce z řad studentů středních škol nebo laiků, než pro odborníky, kteří mají pro takovéto vyhledávání k dispozici specializované nástroje přímo v chemických databázích. Jeden z nich představím v příslušné podkapitole.

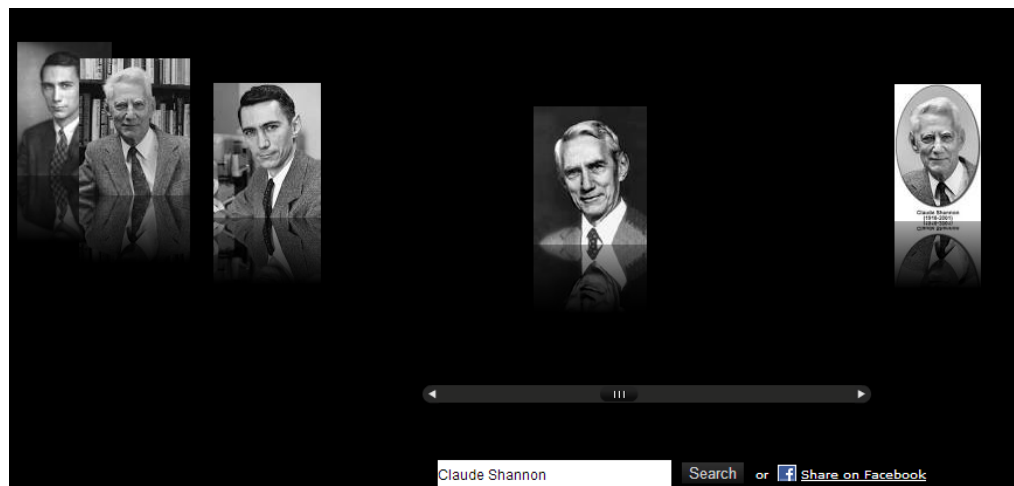
### **3.1.2 Vyhledávání tváří**

Rozpoznávání a vyhledávání tváří z obrazových informací se uplatňuje například v kriminalistice a forenzních vědách, ale také na sociálních sítích nebo ve vyhledávání na volném internetu. Rozpoznávání tváří funguje na základě porovnávání charakteristik obličeje, jako jsou vzdálenosti mezi očima, jejich tvar, šířka a tvar tváře, posazení a tvar nosu apod., s nějakou vnitřní bází obrázků již identifikovaných osob [FACIAL RECOGNITION SYSTEM, 2001].

Na sociálních sítích, jakou je např. *Facebook*, se používá rozpoznávání tváří na fotografiích, které uživatelé na síti sdílí. Funguje jednak tak, že na obrázku dokáže rozlišit lidské tváře od jiných zachycených objektů. Systém rozpoznávání tváří na

*Facebooku* ale dokáže také nalezené tváře na obrázcích spojit s konkrétní osobou, ovšem porovnává tváře jen s osobami z okruhu přátel vlastníka fotografie, nikoliv s tvářemi všech uživatelů. Pokud systém někoho rozpozná, navrhone vlastníkovi označení dané osoby na fotografii [FACEBOOK, 2014]. Samotné vyhledávání podle tváří *Facebook* nepodporuje – možnost porovnat nějakou svou fotografii například s profilovými fotografiemi uživatelů sítě a zjistit tak, kdo se na ní nachází, by rozhodně byla narušením soukromí, a to i přesto, že profilové fotografie jsou veřejné.

Na volném internetu jde tváře vyhledávat například pomocí nástroje *Facesærch* (<http://www.facesærch.com/>), který využívá index obrázků vyhledávače *Google* [About]. Vyhledávání pak probíhá stejně jako v *Google Image Search*, tedy podle popisných informací. Služba velmi hezky prezentuje výsledky jako galerii portrétů, kterou lze listovat pomocí posuvníku (obr. č. 6).



**Obr. č. 6: Zobrazení výsledků ve vyhledávači Facesærch**

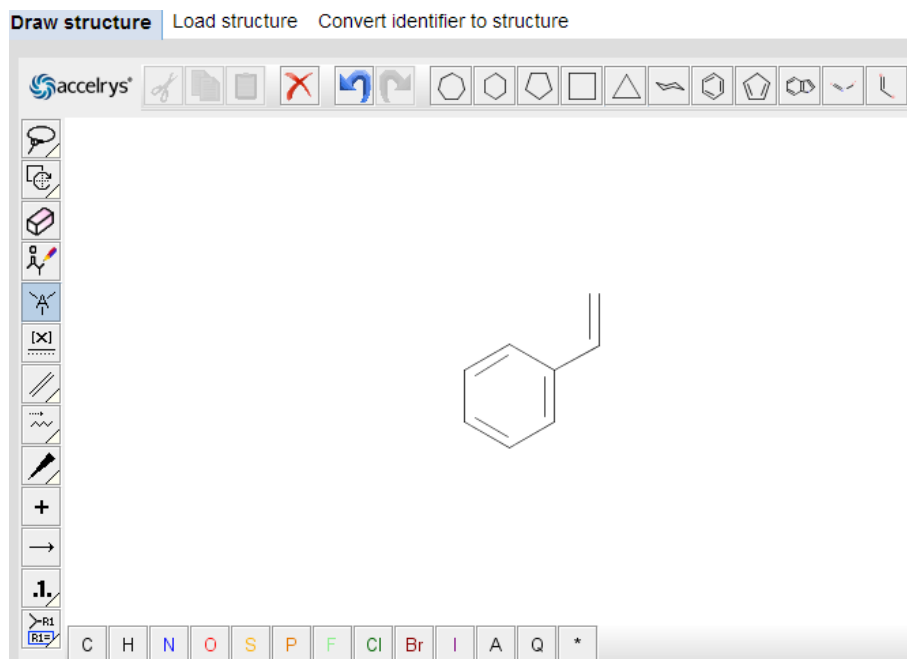
Vyhledávání tváří v oblasti forenzních věd funguje na principu porovnávání obličejových charakteristik ze vstupního obrázku s indexem charakteristik nějaké své interní báze. Na základě shody mezi vstupními daty a daty v bázi se pak snaží identifikovat osobu na vstupním obrázku. Vzhledem k tomu, že identifikace podle tváře probíhá proto, že daná osoba je podezřelá ze spáchání trestného činu, je potřeba vyvinout co největší snahu k nalezení shody. Je zde možnost, že daná osoba nějakým způsobem změnila svou vizáž, ať už záměrně (vousy, brýle apod.), nebo přirozeně (stárnutí), a s tím je třeba počítat. To je první problém, se kterými se vyhledávání podle tváří v této oblasti potýká. Dalším je nutnost porovnávat s bází nejen fotografie, ale také portréty sestavené dle popisu třetích osob. Zároveň by vyhledávání kromě

klasických obličejových charakteristik mělo být schopné zohlednit i speciální znaky jako jizvy, tetování, pihy a znaménka. Faktem zůstává, že i když tyto systémy alespoň nějakým způsobem zmíněné problémy řeší, závěrečný krok – interpretaci vyhledaných výsledků – musí stále provádět člověk [JAIN, 2012].

### 3.1.3 Vyhledávání chemických vzorců

Zvláštním druhem vyhledávání obrázků kreslením je vyhledávání strukturních chemických vzorců. Vzhledem k tomu, že se jedná o využití tohoto typu vyhledávání ve specializovaném prostředí, může nástroj obsahovat širokou škálu funkcí pro snazší konstrukci rešeršních dotazů odpovídajících konvencím příslušného oboru, v tomto případě tedy chemii.

Pro účely testu tohoto nástroje jsem zvolila databázi *ChemSpider* (<http://www.chemspider.com/>), která je volně dostupná a možností vyhledávat podle strukturních chemických vzorců disponuje (obr. č. 7). Její nástroj používá software *Java*, takže pro použití nástroje je třeba mít tento software v počítači nainstalovaný.

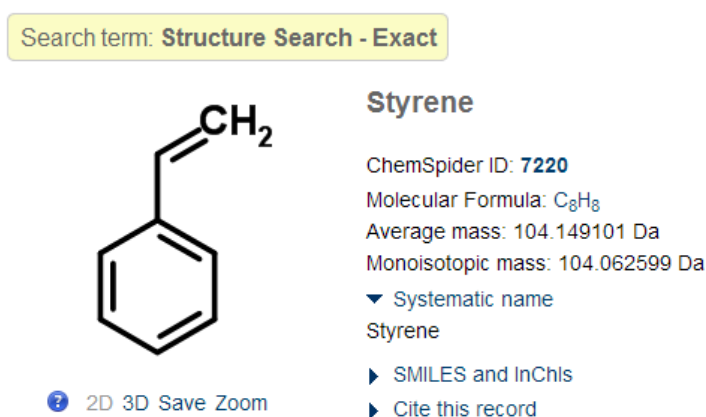


**Obr. č. 7: Rozhraní nástroje databáze *ChemSpider* pro kreslení strukturních vzorců s nakreslenou molekulou**

Specializovanými funkcemi nástroje jsou například možnost přímo kreslit benzenová jádra a jiné cyklické sloučeniny, vytvářet z nich sítě (např. hexové), snadno zdvojit a ztrojit vazby, vybírat z reálných chemických prvků (tj. není možné vytvořit vazbu mezi náhodnými písmennými zkratkami) a mnoho dalších. Po

ukončení kreslení lze ověřit formální správnost vzorce kliknutím na volbu „Clean molecule“, čímž dojde k vyčištění obrázku od chyb např. ve valenci prvků (valence prvků je v chemii daná, je to maximální počet vazeb, které z daného atomu mohou vycházet. Např. uhlík je čtyřvalenční, nemůže tedy nikdy vytvořit více vazeb). Je také možné zkonstruovat rešeršní dotaz nahráním vlastního obrázku. Po vyhledání se zobrazí jako výsledek strukturní vzorec dané molekuly tak, jak je veden v databázi, a další informace o této látce (obr. č. 8).

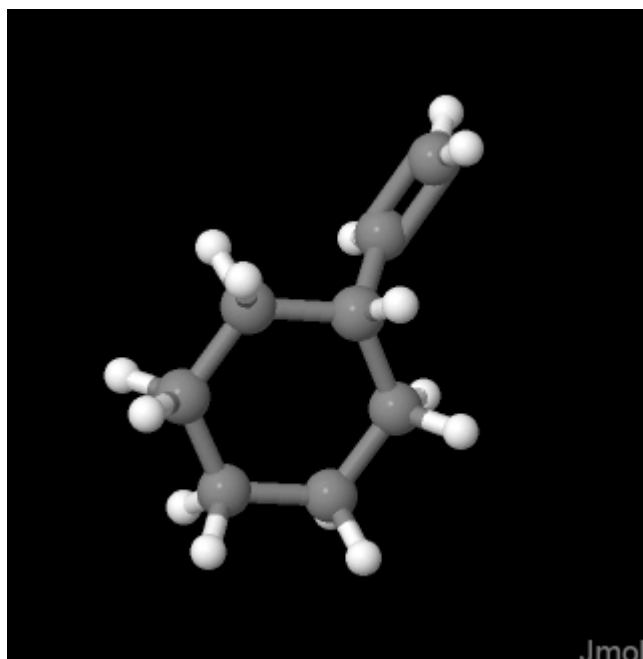
Nástroj po vyhledání dovede převést obrázek na barevný trojrozměrný model molekuly, kterým se dá libovolně otáčet. Na obr. č. 9 je styren, který jsem použila jako testovací vstup pro vyzkoušení nástroje.



**Obr. č. 8: Výsledek vyhledávání v databázi *ChemSpider* podle strukturního vzorce s přesnou shodou**

Toto vyhledávání je důležité proto, že chemické látky se mohou shodovat v sumárním vzorci, který pouze shrnuje druh a počet atomů v molekule, ale jejich vnitřní struktura je různá a jedná se o naprosto odlišné sloučeniny. Například výše zmíněný styren se sumárně zapíše jako  $C_8H_8$ , což může představovat ještě třeba benzocyklobutan, ale i další látky.





Obr. č. 9: 3D model styrenu v databázi *ChemSpider*

### 3.2 Vyhledávání hudebních informací

Stejně jako u vyhledávání obrázků se nejdříve uplatnil slovní popis, tak i zvukové dokumenty lze vyhledávat podle popisných údajů a metadat – podle autora a názvu skladby, použitých hudebních nástrojů nebo žánru, délky, stáří a místa nahrávky aj., ale v některých internetových službách třeba i podle uživatelských tagů označujících například náladu skladby nebo činnost, ke které se její poslech hodí. Takových tagů využívá například služba *8tracks* (<http://8tracks.com/>), kde je podle nich možné vyhledávat celé playlisty sestavené registrovanými uživateli.

Další možností vyhledávání zvukových dokumentů, které je z pohledu inteligentního vyhledávání mnohem zajímavější, je podle obsahu, někdy nazývané vyhledávání hudbou. Vyhledávání hudbou může být realizováno několika způsoby [VOCŮ, 2012]:

- Za použití mikrofону: vyhledávání zpěvem nebo jiným hlasovým projevem, vyhledávání vytřukáváním.
- Za použití nějakého nástroje (míněno programového nástroje): vyhledávání notovým zápisem, vyhledávání zahráním, vyhledávání obrysem melodie.
- Za použití hudební nahrávky: vyhledávání ukázkou.

Kromě vyhledávání ukázkou poskytuje všechny způsoby portál *Musipedia* (<http://www.musipedia.org/>), otevřená hudební encyklopedie.

Vyhledávání hudbou je užitečné z jednoho prostého důvodu – lze ho uskutečnit, i když neznáme jedinou popisnou informaci o hledané skladbě, což se v případě hudby stává docela často. Nezapamatovali jsme si slova, nebo je skladba ani nemá, typicky třeba filmová hudba. Není nic snazšího než ji zabroukat do mikrofону. Podobně vyhledávání vyťukáváním, které se ale soustředí na rytmus skladby, nikoli na melodii.

Při vyhledávání hlasovým projevem *Musipedia* nahraje ukázkou, kterou uživatel předvede, a porovná ji s databází. V případě vyhledávání vyťukáváním zase využívá klávesnici, na níž uživatel poklepáváním na libovolnou klávesu znázorní hledaný rytmus.

Vyhledávání zahráním lze realizovat buď připojením externího pianu k počítači, nebo přepnout na nástroj, který používá pro konstrukci dotazu klávesnici (obr. č. 10). Ovládání je jednoduché a intuitivní, vytvořenou melodii je možné si před spuštěním vyhledání přehrávat. Pokud by někomu přesto manipulace s klávesnicí coby s pianem dělala potíže, nabízí *Musipedia* ještě třetí možnost – flashovou aplikaci simulující piano, která se ovládá pomocí myši a zní opravdu jako klavír (obr. č. 11). Hledanou shodu s databází systém směřuje buď na úplný začátek skladby, nebo někde poblíž něj, případně ji není třeba vůbec omezovat a nechat porovnávat zahraniou melodii s kompletními skladbami. Dotaz lze vždy doplnit ještě slovním popisem.

Nástroj pro vyhledávání obrysem melodie slouží pro ty, kdo nedokáží přesně zrekonstruovat melodii podle tónů, ale pouze určit, jestli je tón oproti tomu předcházejícímu výše, níže, nebo se nezměnil. Takto zkonstruovaný dotaz se skládá z prvního pevně daného tónu, který určuje začátek melodie, a z množiny prvků reprezentujících další tóny, která se skládá ze tří opakujících se instancí. Tyto instance reprezentují zmíněný vztah k předcházejícímu tónu. Dotaz lze ještě doplnit nějakou slovní frází.

Active keys (how to use the keyboard):

Your melody:

e<sup>4</sup> f<sup>4</sup> e<sup>4</sup> d<sup>4</sup> c<sup>4</sup> d<sup>4</sup> c<sup>2</sup> c<sup>4</sup> f<sup>4</sup> f<sup>4</sup> g<sup>4</sup> a<sup>4</sup> g<sup>2</sup>

Delete note ["x" key]  Use flat signs instead of sharp signs

Play

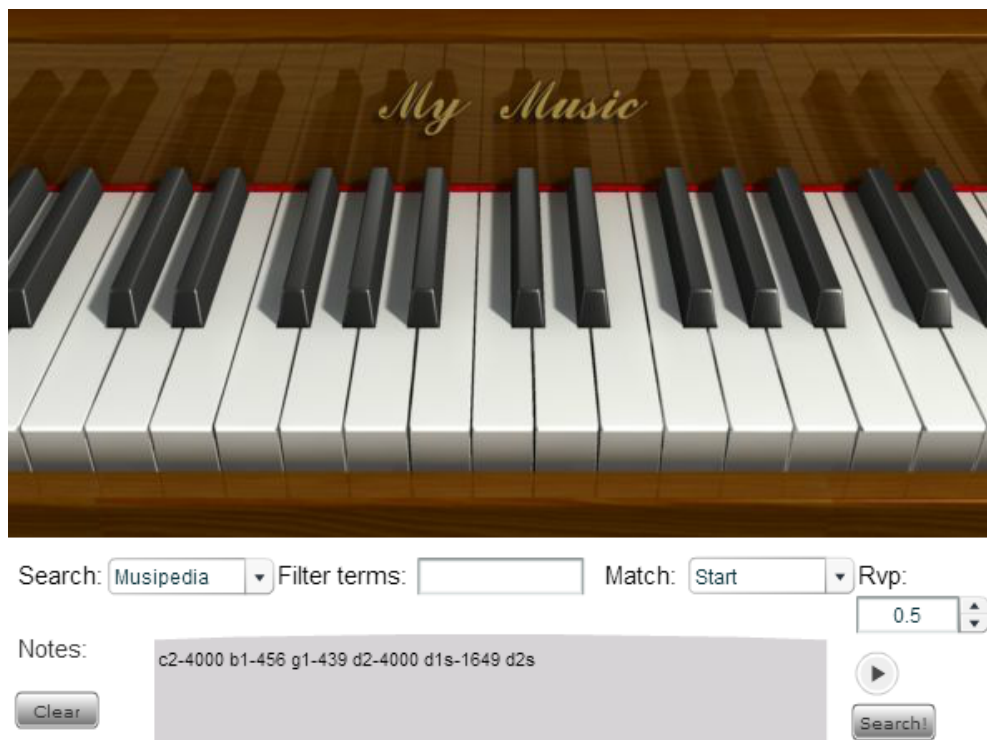
Keywords and/or "search phrases":

Look for matches

**Obr. č. 10: Nástroj pro vyhledávání zahráním pomocí počítačové klávesnice v online hudební encyklopedii *Musipedia***

Vyhledávání podle popisných informací o zvukových dokumentech se uplatňuje jako hlavní způsob vyhledávání v komerčních hudebních knihovnách, kde si lze vyhledané skladby rovnou zakoupit (např. *iTunes*), nebo v databázích nehudebních zvukových souborů, kde by vyhledávání napodobením zvuků ani nebylo možné. Příkladem takové databáze je *Freesound* (<https://www.freesound.org/>), kde si po přihlášení lze stáhnout například zvuk kroků v lese nebo rozbíjeného skla.

Vyhledávání ukázkou *Musipedia* bohužel nepodporuje. Tento typ vyhledávání se uplatňuje spíše u mobilních aplikací, které umožňují instatním poslechem hudby určit, o jakou skladbu se jedná. Dosahují toho vytvořením akustického otisku, což je soubor charakteristik, které program z ukázky získá (frekvence tónů, rychlost skladby na základě četnosti rytmických úderů apod.) [ACOUSTIC FINGERPRINT, 1981]. *Soundhound* nebo *Shazam* jsou příklady takových aplikací.



**Obr. č. 11: Nástroj pro vyhledávání zahráním pomocí flashové simulace piana v online hudební encyklopedii *Musipedia***

### 3.3 Vyhledávání audiovizuálních informací

Jestliže vyhledávání obrazových a zvukových informací nelze realizovat úplně snadno, logicky vyhledávání informací složených kombinací předchozích dvou bude ještě náročnější. V této oblasti se uplatňuje v podstatě jen klasické vyhledávání podle popisných informací, klíčových slov a tagů a následné filtrování výsledků podle zpřesňujících charakteristik. Vyhledávání podle obsahu zatím realizovat neumíme a upřímně řečeno já si neumím ani moc představit, jak by konstrukce takového videodotazu probíhala. Zaznamenávat video i se zvukem dokážeme, ale jak pomocí takových nástrojů vytvořit dotaz, podle kterého by vyhledávač našel třeba film, to je tedy otázka. Šlo by to snad jedině podle již hotové ukázky. Náhradní způsob, jakým se umíme přiblížit vyhledávání videí podle obsahu, je soustředit se na některou z jeho složek – audiální nebo vizuální.

Je obecně známo, že vyhledávání ve vizuální části videí zatím nefunguje dobře, a tak se používá jen ve velmi lehké formě. Např. se vizuální část analyzuje pro rozpoznání lidí od neživých předmětů apod. Vyhledávače se soustředí na indexaci audiální části, ze které vytváří transkripcce, případně indexují textové části videa, pokud se ve videu vyskytují [VIDEO SEARCH ENGINE, 2001]

Z toho vyplývá, že se v tomto vyhledávání mohou uplatnit některé inteligentní prvky – jsou jimi rozpoznávání řeči a textu. Vyhledávače, které tyto služby podporují, jsou schopny na základě slovního dotazu určit, kde přesně ve videu se hledaná fráze nachází, a kterou část videa tedy má smysl si pustit k naplnění dané informační potřeby. Příkladem takového vyhledávače je *TalkMiner* (<http://talkminer.com/>), který vyhledává ve vlastní bázi záznamů z přednášek a prezentací.

Jedním z nejnovějších vyhledávačů v této oblasti je *Thinkglue* (<http://www.thinkglue.com/searchdemo.aspx> – zatím ve verzi beta), který využívá technologie automatického rozpoznávání řeči, obrazové analýzy (identifikace klíčových objektů ve videu) nebo text miningu (zpracování přirozeného jazyka – rozpoznání např. jmen osob nebo míst).

Nejznámější službou zabývající se videodokumenty je bez pochyby Youtube (<https://www.youtube.com/>), z dalších můžeme uvést například Dailymotion (<http://www.dailymotion.com/>). Tyto portály slouží také pro sdílení videí, uživatelé je sem mohou sami nahrávat. Pro lepší orientaci ve výsledcích nabízejí mnoho způsobů, jak video filtrovat a řadit. Po přehrání navíc nabízejí uživateli podobná videa, i když podobnost ověřují opět jen na základě klíčových slov nebo tagů.

### 3.4 Enterprise search

Enterprise search (do češtiny by se dalo volně přeložit asi jako vyhledávání podnikových informací) je v některých zásadních bodech podobné internetovému vyhledávání, i když třeba ne v takovém měřítku. Jde především o množství, umístění a různorodost informací. Však také jeho rozvoj souvisel s rozmachem internetu – začaly se ve firmách používat nástroje pro sdílení informací, např. mezi různými odděleními podniku, a ukázalo se, že zpřístupnění informací zaměstnancům s sebou nese řadu výhod (urychlení procesů, které vážly na komunikaci, transparentnost, včasné odhalení chyb atd.). Zároveň to ale přináší potřebu se v nich efektivním způsobem orientovat, což je velmi komplexní problém. Tímto problémem se zabývá enterprise search jakožto odborné odvětví.

Pojem enterprise search lze také chápat jako nástroj nebo řešení pro vyhledávání v korporátních informacích, tedy nejen obecně jako problematiku vyhledávání v tomto prostředí. Theresa Regli ve svém článku parafrázuje prodejce

takovýchto nástrojů: „*enterprise search je odpovědí na potřebu organizace mít jediné, unifikované okénko, které vede ke všemu, co daná organizace zná, a to naprosto kdykoli – dokonalý stroj pro správu znalostí*“ [REGLI, 2008, s. 89]. To je samozřejmě jen ideální představa, takový nástroj neexistuje – naráží nejen na problémy, se kterými se potýká vyhledávání informací obecně, ale i na takové, které jsou dány specificitou podnikového prostředí.

Jestli má totiž takový nástroj fungovat skutečně efektivně pro daný podnik, je záhodno, aby při své „akci“ bral v potaz podnikově-specifické znalosti. Jakožto komplexní nástroj pro správu podnikových informací by měl být potom schopen provádět i další související úkoly, např. data-minig, ideálně přes jedno společné rozhraní bez ohledu na to, o jaký druh informace se jedná nebo kde je uložena.

V současné době nabízí mnoho firem enterprise search řešení. Většinou se jedná o společnosti, které zároveň nabízejí i jiný software a podnikové informační systémy (ERP systémy, ekonomické systémy apod.). Patří mezi ně např. firmy Microsoft, IBM, Oracle nebo SAP. Další alternativou jsou volně dostupné nebo open-source softwary, např. *Apache Solr*.

### **3.4.1 Nalezitelnost**

O nalezitelnosti informací se začalo mluvit v souvislosti s internetem – vzhledem k ohromnému nárůstu množství informací na něm obsažených lze předpokládat, že pokud člověk shání nějakou informaci, s největší pravděpodobností se už na internetu někde vyskytuje. Jenže není k nalezení.

Stejně tak v oblasti podnikových informací – procesy se opakují, problémy, které firma řeší, jsou si často v mnohém podobné. Způsob, jakým byla situace řešena v minulém případě, je ukryt kdesi v hlubinách podnikového intranetu, e-mailové komunikace nebo papírových šanonů, a zabralo by mnohem více času se k němu dostat než řešení prostě vymyslet znovu. To je problém, který se nástroje pro enterprise search snaží vyřešit.

John Mancini [2010] uvádí ve svém článku celkem osm klíčových problémů, se kterými je potřeba se vypořádat pro zlepšení nalezitelnosti dat při implementaci enterprise search řešení:

- 1. Data uložena v mnoha typech souborů** – podnikové soubory byly v průběhu času vytvářeny v mnoha různých systémech a programech, každý je ukládal na jiné místo a v jiném formátu. Ne všechna úložiště jsou přístupná všem zainteresovaným uživatelům a co je důležitější, když už přístupná jsou, soubory lze často otevřít pouze v programu, který je chráněný. Soubory jsou pak nevyužitelné, protože se nelze dostat k jejich obsahu.
- 2. Data „uzamčená“ v systémech** – s růstem podniků stoupá i počet podporovaných aplikací. Potřeba znát všechna místa, kde je možné hledanou informaci najít, a nutnost zacházet s mnoha různými rozhraními značně zpomalují proces nalezení té správné informace a mohou vést ke zmatení.
- 3. Potřeba kategorizace a kontextu** – aplikace, která dokáže uživateli vyhledat dokumenty na základě klíčových slov je užitečná, nicméně schopnost vyhledávače interpretovat text a poskytnout případně související informace a tolik potřebný kontext je nedocenitelná. Pro efektivnější vyhledávání je možnost redefinovat rešeršní dotaz případně specifikovat oblast hledání nutností.
- 4. Opravy překlepů** – běžný problém ve všech rešeršních systémech: nutnost eliminovat chybovost v hláskování. Schopnost systému kontrolovat rešeršní dotaz proti nějakému internímu slovníku často se vyskytujících slov a napovídat uživateli pomocí jednoduché funkce „měli jste na mysli...?“ je důležitou součástí efektivního vyhledávače.
- 5. Lemmatizace** – tvary slov použité v rešeršním dotazu by neměly hrát roli a systém by s nimi měl pracovat jako s různým vyjádřením téhož. Obzvlášť v češtině, která coby flexivní jazyk vyjadřuje gramatické funkce pomocí skloňování, je tato funkcionalita velmi výhodná
- 6. Relevance** – zapojení vážení termínů podle toho, kde se v prohledávaném dokumentu vyskytují, může značně zvýšit šanci, že systém identifikuje dokumenty skutečně relevantní danému rešeršnímu dotazu a zobrazí je ve výsledcích nejvýše.
- 7. Bezpečnost** – ve firemních informacích se nachází mnoho dokumentů, které nejsou přístupné všem, ať už je to z důvodů vnitropodnikové politiky nebo ochrany klienta. Takové dokumenty by vyhledávač měl také indexovat, aby

byly nalezitelné pro osoby oprávněné je prohlížet. Je ovšem třeba respektovat omezení přístupu a obsah zobrazil jen po autentikaci, případně „tajné“ dokumenty ve výsledcích běžným uživatelům vůbec nezobrazovat.

- 8. Dobrá „vyhledávací zkušenost“** – intuitivní a přívětivé uživatelské rozhraní vyhledávače je rovněž velmi důležité. Uživatel musí pokud možno s co nejmenší vynaloženou námahou pochopit, co se po něm chce, a jaké má možnosti, aby to realizoval. Důležitá je proto i zpětná vazba rešeršního systému a rozšířená funkcionalita pro prohlížení výsledků (náhledy apod.)

Pokud je enterprise search engine sám o sobě „silný“, dokáže vylepšení stavu ve výše zmíněných bodech výrazně zkvalitnit jeho výkon.

### **3.4.2 Enterprise search versus webové vyhledávání**

Enterprise search má s webovým vyhledáváním některé společné rysy, a proto se může inspirovat některými technikami, které používají internetové vyhledávače. Stejně jako na webu i ve firemních informacích se vyhledávání potýká se strukturovanými a nestrukturovanými daty, které se prohledávají na základě jednoho dotazu [ENTERPRISE SEARCH, 2014].

Společnost Google například sama nabízí enterprise search řešení, *Google Search Appliance (GSA)*, které v mnohých principech vychází ze zkušeností s internetovým vyhledáváním.

U enterprise search je důležitá možnost customizace řešení, což je dáno různorodými vstupními podmínkami v podnicích, takže je potřeba aby obsahovalo maximální množství nástrojů, díky kterým lze takovéto variability docílit. Některé lze převzít z internetových vyhledávačů (kontrola pravopisu, doplňování dotazů...), jiné bylo potřeba nově navrhnout (relevantní zpětná vazba, náhledy dokumentů...).

Novým způsobem bylo potřeba přistupovat k řazení výsledků – internetový vyhledávač *Google* posuzuje relevanci stránek (mimo jiné) podle page-ranku, tedy množství a významnosti odkazů, které na danou stránku směřují. Tento způsob u enterprise search není možný – vzhledem k tomu, že vyhledávání probíhá uvnitř nějakého intranetu nebo korporátní soukromé sítě, která je chráněna firewallem. *GSA* implementuje relevantní zpětnou vazbu, analyzuje chování a preference uživatelů a učí se tak zlepšovat kvalitu výsledků vyhledávání [GOOGLE, 2001].



V průzkumu poradenské firmy v oblasti informačních technologií Gartner se řešení *GSA* umístilo velmi dobře, nicméně je mu vytýkán jeden podstatný rys související s vyhledáváním na internetu – nemožnost určit, proč na daný dotaz vyhledávač vrátil právě takový balík výsledků. Společnost Google totiž považuje své relevantní modely za součást firemního tajemství a významnou konkurenční výhodu, takže uživatel ani nemůže přesně vědět, proč na daný dotaz dostal právě takové výsledky [ANDREWS, 2013].

Enterprise search řešení dále musí umět zacházet jak s tradičními tak s novými zdroji dat. Mezi ty tradiční patří datové servery, softwary pro správu podnikového obsahu nebo e-mailové archivy. Novinkou pak jsou cloudová úložiště pro uchovávání a sdílení dat jako služby *Dropbox*, *Google Drive* nebo *Trello*.

Oblast enterprise search otevřela zcela nové možnosti vývoje ve vyhledávání informací, protože svými specifickými potřebami vytvořila příležitost uplatnit nové postupy a nástroje, a posunout tak vyhledávání informací zase o něco dál.

## **4 Směry vývoje vyhledávání informací v uzavřených informačních systémech a na volném internetu**

Obecně lze k této problematice říci, že na volném internetu se prvky inteligentního vyhledávání implementují do vyhledávačů rychleji, než v uzavřených systémech. Je to dáno jednak podmínkami v obou prostředích, ale také třeba uživatelskou základnou.

Pokud jde o podmínky, ve kterých vyhledávače fungují, tak na volném internetu jsou k integrování inteligentních prvků v podstatě „tlačeny“. Způsobuje to rychlý a masivní nárůst množství informací a různorodost jejich formátů, na což musí vyhledávače pružně reagovat. Uzavřené systémy se často specializují, takže množství informací v nich nenarůstá tak rapidně, a pokud jde o různorodost formátů, systémy si ji sami korigují – pokud zahrnují i netextové dokumenty, vyhledávají se většinou tradičně podle slovních popisů.

Okruh uživatelů má na implementaci prvků inteligentního vyhledávání také nemalý vliv. Pokud uživatel na internetu nemůže požadovanou informaci v daném vyhledávači najít, prostě zkusí nějaký jiný. Vývojáři vyhledávačů se proto snaží je neustále vylepšovat a inovovat, aby dokázali uživatelským potřebám co nejlépe vyhovět a uživatele si udržet. Zároveň integrací uživatelsky „zajímavých“ prvků mohou přilákat uživatele nové, i kdyby to mělo znamenat, že uživatelé použijí vyhledávač jen proto, aby si takové vyhledávání vyzkoušeli (např. vyhledávání hlasem nebo kreslením). Uzavřené systémy mají uživatelskou základnu mnohem stabilnější, takže nemusejí na technologické novinky reagovat tak rychle, nicméně i zde existuje konkurenční prostředí a „být první“ se může vyplatit. Na druhou stranu uživatelé uzavřených systémů jsou řekněme konzervativnější, takže určitá opatrnost v implementaci nových prvků a radikálních proměn rozhraní je zde na místě.

### **4.1 Uživatelské rozhraní a nápověda**

Nejjednodušším způsobem jak zapojit nějaký inteligentní prvek do vyhledávání je modifikovat uživatelské rozhraní na základě studií uživatelského chování a poznatků z HCI, a vytvořit propracovanou a sofistikovanou nápovědu.

Pro uživatelské rozhraní na volném internetu rozhodně platí „čím jednodušší, tím lepší“. Když se podíváme na základní rozhraní *Googlu*, opravdu už není moc co ubrat – máme jedno vstupní okénko, všechny ostatní nástroje, které jsou uživateli využívány ve výrazně menším měřítku než základní vyhledávání, jsou skryty. Pokročilejší uživatel se k nim musí proklikat sám – dá se předpokládat, že takový uživatel už nějakou zkušenost s vyhledáváním má a bude tedy vědět, jak na to. Vyhledávače na volném internetu si takto prosté rozhraní mohou dovolit – dotazy, které do nich uživatelé zadávají, totiž zahrnují široké spektrum informačních potřeb a týkat se mohou naprosto čehokoli. Vytvářet speciální rozhraní, které by počítalo se všemi možnostmi, by bylo nesmyslně složité a nakonec nutně uživatelsky nepřívětivé. Já osobně tento koncept „maximální jednoduchosti“ považuji za velmi chytrý – líbí se mi hlavně proto, že pro nenáročného uživatele je naprosto dostačující, a přitom zkušeným uživatelům dává najevo, že vlastně tak trochu počítá s jejich vlastní inteligencí.

V uzavřených systémech je situace složitější. Uživatelské rozhraní musí zohlednit povahu informací obsažených v databázi, obvykle oborové, z níž se dá alespoň v hrubých obrysech vyvodit, jak budou vypadat rešeršní dotazy formulované uživateli (respektive jak vypadat nebudou). Uzavřené systémy tak často konstruují tezaury nebo využívají již existující řízené slovníky pro daný obor. Dalším faktorem je předpoklad, že uživatel přišel nalézt informaci, která je v relativním měřítku velmi konkrétní. Co tím myslím, se nejlépe demonstruje na příkladu: uživatel hledá nejnovější informace o možnostech využití anihilace antihmoty jako zdroje energie. Odpovědí na jeho potřebu může být celý balík relevantních odborných článků nikoli jedna jediná relevantní informace. Přesto se jedná o docela specifický dotaz. Pokud je dotaz úzce zaměřen, pak i když očekává více než jednu relevantní odpověď (spíše bývá žádoucí dostat jich více), je vhodné moci ho i velmi konkrétně zadat do vyhledávače – třeba specifikací jazyků dokumentů nebo doby jejich vydání. Například komerční databáze *Ebsco* používá také docela jednoduché rozhraní s jedním vstupním okénkem, pod kterým se nachází méně výrazné možnosti hledání, kterými lze dotaz omezit. Další doplňující nástroje, kterými jde vyhledávání limitovat nebo zacílit správným směrem, nabízí v druhém kroku – při zobrazení výsledků.

Jak právně nastavit nápovědní systém se ukázalo být docela obtížným. Je to dáno všeobecnou nechtí uživatelů ke studování manuálů a doporučení. Určit hranici mezi

stavem, kdy by nápověda uživatele jen otravovala, a kdy by ji skutečně potřeboval, je neuvěřitelně těžké. Většina uzavřených systémů i webových vyhledávačů raději volí cestu „neotravování“, tedy že kompletní nápověda k funkcím a možnostem vyhledávání je skryta a zobrazí se jen na výslovný pokyn uživatele. Automatickou nápovědu, která se zobrazí sama, systémy používají například jen v opravách pravopisu v rešeršních dotazech, případně v návrzích podobných termínů.

## 4.2 Rozšíření služeb

Cesta, kterou volí uzavřené systémy v oblasti rozšíření služeb, je provázanost na další systémy a služby, například na citační manažery nebo citační databáze, a dále na vytvoření jakéhosi personalizovaného přístupu do databáze – přestože uživatel do systému vstupuje na multilicenci nějaké instituce, např. knihovny, má často možnost si v systému vytvořit osobní schránku na vyhledané fulltexty, ukládat si relevantní záznamy a třídit je do složek, archivovat rešeršní dotazy atd.

Webové vyhledávače také poskytují personalizované vyhledávání ovšem zde je otázka, jestli je pro uživatele přínosem nebo ne. Vyhledávače totiž nezobrazují pro každého uživatele stejné výsledky, ale sledují jeho aktivitu na internetu (historii v prohlížeči) a při vyhledávání tomu prezentované výsledky přizpůsobují. Zjednodušeně řečeno, když dva lidé zadají do webového vyhledávače totožný dotaz, nemusí se jim zobrazit totožné výsledky. Přínos personalizovaného vyhledávání pro uživatele je diskutabilní proto, že ho lze využít v marketingu a elektivněji cílit reklamu.

Webové vyhledávače nabízejí ještě službu bezpečného vyhledávání, která má zbránit zařazení obsahu pouze pro dospělé mezi výsledky (pornografické stránky a obrázky). Bezpečné vyhledávání využívá také zpětnou vazbu uživatelů, kteří mají možnost nahlásit explicitní obsah, který bezpečnostním filtrem prošel.

## 5 Inteligentní vyhledávání jako předmět výzkumu

Výzkum v oblasti inteligentního vyhledávání se soustřeďuje na odborných pracovištích především při univerzitách a ve specializovaných centrech. Vznikají zde odborné články a závěrečné práce zabývající se když ne přímo inteligentním vyhledáváním jako takovým, tak dílčími tématy z tohoto odvětví.

V této kapitole představím nejvýznamnější z těchto pracovišť, která se alespoň v některé části svého zaměření zabývají oblastí inteligentního vyhledávání nebo technologiemi v něm používanými. Konkrétní příklady výstupů výzkumné a studijní činnosti v českém prostředí jsou k nalezení v příložené rešerši ([příloha č. 1](#)).

Představená pracoviště jsou uvedena na základě online průzkumu a vycházejí také z literatury citované v této práci. Jsou zahrnuta také některá významná oddělení výzkumu soukromých firem, jejichž cílem není primárně výzkumná činnost, nýbrž konkrétní výsledky použitelné pro účely podnikání.

### 5.1 Odborná pracoviště a univerzity v ČR

#### České vysoké učení technické

##### Fakulta elektrotechnická, Katedra kybernetiky, Centrum strojového vnímání

Jedná se o výzkumné centrum zabývající se základním a aplikovaným výzkumem počítačového vidění a, robotiky, strojového učení a dalších příbuzných oborů. Za zmínku stojí účast na projektu DIRAC (Detection and Identification of Rare Audio-visual Cues), který se zabýval návrhem metod pro vývoj autonomních umělých kognitivních systémů, které by uměly zachytit, identifikovat a klasifikovat možné rizikové události z audiovizuálních podnětů, čehož by dosahovaly pomocí četných audiovizuálních senzorů [DIRAC, 2013].

##### Fakulta informačních technologií, Software Engineering Group

Tato pracovní skupina se zabývá databázemi, XML technologiemi a vývojem programovacích jazyků a jejich implementací. V současné době zpracovává velmi specifické projekty, které souvisí především s programovacími jazyky.

### Fakulta informačních technologií, Katedra počítačových systémů

Na této fakultě se problematice vyhledávání informací věnují ještě na Katedře počítačových systémů – konkrétně sémantizaci webu a výzkumu metod vyhledávání na webu s využitím sémantiky vyhledávané informace, distribuovanému dobývání znalostí, výzkumu inteligentních a heuristických postupů pro podporu vyhledávání informace ve webovém prostoru a XML technologiím [KATEDRA POČÍTAČOVÝCH SYSTÉMŮ, 2013].

### **Masarykova univerzita**

#### Filozofická fakulta, Kabinet informačních studií a knihovnictví

Na brněnské Masarykově univerzitě se informační studia a knihovnictví vyučují na KISK. Zaměření je podobné jako na pražském ÚISK, obor je koncipován jako „*transdisciplinární studium s propojením na Fakultu informatiky a Fakultu sociálních věd*“ [KABINET INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ, 2004].

Jedním z projektů, na kterých se KISK podílel, byl alternativní elektronický katalog *Beth* (2008), jehož cílem bylo zapojit do vytváření, úprav a třídění záznamů své uživatele, umožnit jim vzájemnou komunikaci a sdílení názorů s ostatními, zkrátka implementovat některé prvky sociálních sítí. Katalog byl zamýšlen jako experimentální, po několika letech byl jeho provoz ukončen především kvůli nízké využívanosti, ale i z dalších, technických důvodů [BETH – ROZŠÍŘENÝ ONLINE KATALOG, 2008].

#### Fakulta informatiky

Na této fakultě se problematice vyhledávání informací věnují obory magisterského studijního programu, např. *Informační systémy* a *Počítačové systémy*, nicméně konkrétně problematikou inteligentního vyhledávání se alespoň zčásti zabývá obor *Umělá inteligence a zpracování přirozeného jazyka*, který se kromě jiného zaměřuje na zpracování přirozeného jazyka, reprezentaci a management znalostí, strojové učení nebo dolování z dat [FAKULTA INFORMATIKY MUNI, 1996].

Na fakultě také funguje Centrum zpracování přirozeného jazyka, jehož cílem je teoretický i praktický výzkum v oblasti korpusové lingvistiky, lexikálních databází,

reprezentace znalostí, reprezentace významu výrazů přirozeného jazyka a využití metod strojového učení pro desambiguaci korpusových dat. Centrum se podílí na vývoji nástrojů ve zmíněných odborných oblastech, např. na *Internetové jazykové příručce Ústavu pro jazyk český Akademie věd ČR* (<http://prirucka.ujc.cas.cz/>), nástroji *JusText* (<http://nlp.fi.muni.cz/projekty/justext/>), který po zadání URL do vstupního okna eliminuje z daného webu všechny netextové části, a mnohých dalších.

Z hlediska výzkumu v oblasti inteligentního vyhledávání jsou zajímavé ještě Laboratoř dobývání znalostí a především Laboratoř DISA – Data Intensive Systems and Applications, která se zabývá pokročilými metodami vyhledávání v digitálních datech. Činnost této laboratoře se soustřeďuje kolem projektu *MUFIN – Multi-feature Indexing Network* (<http://mufin.fi.muni.cz/>), který je podrobněji popsán v podkapitole o vyhledávání obrazových informací.

### **Mendelova univerzita v Brně, Provozně ekonomická fakulta, Ústav informatiky**

Pracovní skupina s názvem Inteligentní systémy a multimédia se na tomto ústavu věnuje výzkumu a výuce v oblastech zpracování obrazu, komplexních mobilních aplikací, geografických informačních systémů nebo vývoje softwaru pro aplikace matematických ploch. Má k dispozici specializované laboratoře virtuální reality, počítačové grafiky a inteligentních systémů [PROVOZNĚ EKONOMICKÁ FAKULTA, 2014]

### **Univerzita Karlova v Praze**

#### Filozofická fakulta, Ústav informačních studií a knihovnictví

Vyhledávání informací jakožto součást informační vědy patří mezi důležité oblasti vyučované na ÚISK. Informační studia a knihovnictví jsou nicméně humanitním oborem a jako takové se soustřeďují na praxi ve vyhledávání, na pořádání znalostí, rešeršní strategie a celkový přehled o databázích, digitálních knihovnách a možnostech a směrech vývoje v této oblasti spíše než na technické a programové provedení.

Vyhledáváním informací se zabývají studenti ve svých závěrečných pracích, a to především absolventi informačního zaměření navazujícího magisterského studia,

nicméně jsou zde také realizovány odborné projekty, např. *Archivace a zpřístupnění digitálních verzí kvalifikačních prací na UK v Praze (2005-2007)*, *Výzkum a implementace sémantické analýzy textu ve webovém prostředí (2006)* [ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ, 2014].

#### Matematicko-fyzikální fakulta, Katedra softwarového inženýrství

Na rozdíl od ÚISK se Katedra softwarového inženýrství MFF soustřeďuje právě technickou a programovou částí vyhledávacího procesu, vyhledávací algoritmy a matematické modely vyhledávání (kromě jiného). Na katedře funguje od roku 2006 výzkumná skupina SIRET (SIMilarity RETrieval), která se zabývá metodami pro efektivní a výkonné podobnostní vyhledávání v databázích nestrukturovaných objektů. Jejím hlavním zájmem jsou tři oblasti: metody indexování podobnosti, biologické aplikace podobnostního vyhledávání a indexování obrazových databází pro vyhledávání založené na obsahu [SIRET RESEARCH GROUP, 2014].

Katedra se také podílí na mnoha výzkumných projektech, účastnila i se projektů Ústavu informatiky Akademie věd ČR ukončených v letech 2008 a 2009 (zmíněny níže). V současnosti je v realizaci např. grant *Podobnostní nemetrické vyhledávání v rozsáhlých komplexních databázích* vypsany Grantovou agenturou České republiky.

#### Matematicko-fyzikální fakulta, Ústav formální a aplikované lingvistiky

Tento ústav se zaměřuje na zpracování textu, strojový překlad, zpracování přirozeného jazyka, sémantiku a další obory. Ústav také zpracovává výzkumné a grantové projekty, v současné době například probíhají projekty *Interactive information retrieval in audiovisual dialogue corpora* a *KHRESMOI – medical information analysis and retrieval*. První jmenovaný se zabývá vyhledáváním v audiovizuálním korpusu, přičemž důraz je kladen na interaktivní metody vyhledávání informací. Cílem druhého jmenovaného projektu je vyvinout vícejazyčné multimodální vyhledávací systém pro biomedicínské informace a dokumenty [INSTITUTE OF FORMAL AND APPLIED LINGUISTICS, 2014].

### **Ústav informatiky Akademie věd České republiky**



Jedná se o významnou veřejnou instituci neuniverzitního typu, která si klade za cíl zaprvé „provádět výzkum v oblasti informatiky v souladu s aktuálními trendy v teorii a aplikacích informatiky a tyto trendy také spoluvytvářet“ a zadruhé „přenášet získané znalosti do společnosti, konkrétně vědecké komunitě, studentům a aplikačním oblastem vědy a výzkumu“ [ÚSTAV INFORMATIKY AV ČR, 2013]. Nejvýznamnější grantové projekty související s inteligentním vyhledáváním informací, které Ústav zpracovával, jsou následující (v závorce rok ukončení projektu):

- *Metody umělé inteligence v GIS* (2011)
- *Inteligentní softwaroví zprostředkující agenti pro mediaci nekompatibilních sémantických Webových Služeb* (2010)
- *Metody inteligentních systémů při dobývání znalostí a zpracování přirozeného jazyka* (2009)
- *Inteligentní modely, algoritmy, metody a nástroje pro vytváření sémantického webu* (2008)
- *Inteligentní vyhledávání v dokumentografických systémech* (2002)
- *Vyhledávání v textových databázích s využitím principů umělé inteligence a neuronových sítí* (1996)
- *Tvorba inteligentních systémů* (1995)

### **Ústav teorie informace a automatizace Akademie věd ČR**

Tento ústav je dalším zástupcem výzkumných neuniverzitních institucí. Z oblasti inteligentního vyhledávání provádí výzkum na poli umělé inteligence, zpracování signálů a obrazu a rozpoznávání obrazu. Pracoviště zároveň spolupracuje s UK, ČVUT a VŠE zajištěním výuky některých předmětů a vedením doktorandů.

Ústav se dělí na jednotlivá odborná oddělení, která specifikují svůj zájem na konkrétní oblast výzkumu. Oddělení Zpracování obrazové informace provádí výzkumy, jejichž výsledky nalézají různá uplatnění, např. v rekonstrukci nekvalitních obrázků (snímky z bezpečnostních kamer apod.) nebo medicínském snímkování.

Zajímavým projektem je databáze LEAF, která dokáže rozpoznat druh dřeviny podle obrázku jejího listu [TREE LEAF DATABASE MEW2010, 2010].

## **Vysoká škola ekonomická, Fakulta informatiky a statistiky**

### Katedra informačního a znalostního inženýrství

„*Informační a znalostní inženýrství je chápáno jako soubor metod zpracování (získávání, reprezentace, transferu a využívání) informací a znalostí*“ [KATEDRA INFORMAČNÍHO A ZNALOSTNÍHO INŽENÝRSTVÍ, 2000]. Předmětem výuky a výzkumu na katedře jsou například automatické získávání a dobývání znalostí, metody inteligentního vyhledávání v informačních zdrojích včetně internetu nebo zpracování přirozeného jazyka. Nejvíce se na zmíněné problematice soustředí navazující magisterské studium se studijní specializací *Znalostní technologie*.

Na katedře také působí výzkumná skupina Semantic Web and Ontological Engineering (SWOE), jejímž hlavním předmětem výzkumu je zpracování ontologií včetně vizualizace a problematika linked data.

### Katedra informačních technologií

Obor vyučovaný na magisterském stupni Kognitivní informatika nejvíce koresponduje s problematikou inteligentního vyhledávání. Jedná se o hraniční obor využívající poznatky z informatiky, softwarového inženýrství, umělé inteligence, kybernetiky, kognitivních věd, filosofie, lingvistiky a dalších věd [KOGNITIVNÍ INFORMATIKA, 2000].

## **Vysoké učení technické v Brně, Fakulta informačních technologií**

Problematika související s inteligentním vyhledáváním je vyučována na Ústavu informačních systémů a Ústavu inteligentních systémů, nicméně vyučované předměty jsou více zaměřeny na praktickou informatiku a programování. Naproti tomu Ústav počítačové grafiky a multimédií se zabývá několika oblastmi souvisejícími s inteligentním vyhledáváním: rozpoznávání řeči, dolování informací z řeči, tvorba řečových databází.

Na této fakultě jsou v činnosti také četné výzkumné skupiny, například Výzkumná skupina inteligentních systémů, která se věnuje teoretickému a praktickému výzkumu v oblastech umělé inteligence a soft-computingu, programových agentů a multiagentních systémů, robotických systémů a inteligentních senzorů. Další výzkumné skupiny, které se rovněž zabývají umělou inteligencí a technologiemi souvisejícími s inteligentním vyhledáváním jsou Výzkumná skupina znalostních technologií a Výzkumná skupina zpracování obrazu a videa. Tyto skupiny kromě výzkumu ještě zajišťují výuku tematicky odpovídajících předmětů na celé fakultě a studenti se podílejí na jejich činnosti psaním závěrečných prací [FAKULTA INFORMAČNÍCH TECHNOLOGIÍ VUT V BRNĚ, 2014].

### **Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra kybernetiky**

Na magisterském stupni je zde možná studovat specializaci Umělá inteligence, která se mimo jiné zabývá aplikací inteligentních metod rozhodování, navrhováním a realizací systémů komunikace člověk-počítač mluvenou řečí, počítačovým vidění, neuronovými sítěmi, znalostními systémy apod.

Výuku zajišťuje Oddělení umělé inteligence, které zároveň zapojuje studenty do řešení grantových a vývojových projektů. Hlavním zaměřením skupiny je hlasová komunikace člověka s počítačem v přirozeném jazyce – jedním z výstupů činnosti skupiny je výzkum vyhledávání informací v řečových datech. Vyhledávání jednotlivých slov či krátkých frází již skupina uplatnila i prakticky, aplikaci v audiovizuálním archivu svědků Holokaustu (<http://sfi.usc.edu/collections/holocaust>), tedy v jeho česky mluvené části [KATEDRA KYBERNETIKY ZČU, 2014].

## **5.2 Zahraniční odborná pracoviště**

Zahraničních odborných pracovišť a univerzit, kde se problematikou inteligentního vyhledávání alespoň částečně zabývají, je opravdu hodně. Vybrala jsem proto pouze ta pracoviště, kde je tato oblast hlavním předmětem zkoumání, případně ta, která jsou podle mě významná.

### **CIIR – Center for Intelligent Information Retrieval (School of Computer Science, University of Massachusetts)**

CIIR se zabývá zkoumáním prakticky všech oblastí a technologií, které by mohly nějakým způsobem přispět k vývoji a zefektivnění vyhledávání informací. Pracovníci centra každoročně publikují desítky odborných článků a konferenčních příspěvků, centrum navíc spolupracuje s vládními a průmyslovými organizacemi a podílí se na vývoji a podpoře open-source softwaru [CIIR – CENTER FOR INTELLIGENT INFORMATION RETRIEVAL, 2014]. Ředitelem centra je W. Bruce Croft, který je uznávanou kapacitou v oboru a autorem mnoha odborných článků a knih, z nichž některé sloužily jako odborné podklady pro tvorbu této práce, jsou v ní citované a uvedené v seznamu literatury.

### **SIGIR – Special Interest Group on Information Retrieval (Association for Computing Machinery)**

Výzkumná skupina SIGIR se soustřeďuje na všechny aspekty ukládání, vyhledávání a distribuce informací včetně rešeršních strategií nebo evaluací systémů [SIGIR, 2014]. Skupina se také podílí na pořádání odborných konferencí a vydává vlastní newsletter *SIG-IRlist*, který je možné zdarma odebírat přes e-mail. SIGIR vydává ještě *The SIGIR Forum* pro účastníky konferencí, které se skládá ze dvou půlročních newsletterů a jednoho zvláštního vydání konferenčních příspěvků z akcí pořádaných skupinou.

### **IEEE – Institute of Electrical and Electronics Engineers**

IEEE se sice nesoustředí na vyhledávání informací coby svůj ústřední zájem, nicméně považují za vhodné zde tuto asociaci uvést, protože se jedná o „světově největší profesionální asociaci pro pokrok v technologii“ [IEEE, 2014]. Asociace se skládá z ohromného množství pracovních skupin zabývajících se jednotlivými obory. IEEE se výrazným způsobem podílí na podpoře vědy a výzkumu v oblasti informačních technologií pořádáním konferencí, publikováním a v neposlední řadě tvorbou a správou databáze *IEEE Xplore*, která pokrývá veškerou publikační činnost IEEE a obsahuje tak více než 3 miliony fulltextových dokumentů [IEEE, 2014a].

Databázi je možné volně prohledávat, nicméně k fulltextům lze získat přístup až po zakoupení licence.

### **The Information Retrieval Group (IBM)**

Výzkumná skupina soukromé firmy IBM se soustřeďuje na problematiku vyhledávání nestrukturovaných informací, jejich reprezentaci, ukládání a organizaci. Mezi zkoumané oblasti patří například XML vyhledávání nebo enterprise search. Cílem skupiny je hledat a následně zkonstruovat řešení, která se poté stanou součástí informačních systémů a dalších produktů firmy IBM. Členové skupiny přispívají i do akademické komunity publikováním odborných článků, aktivní účastí na konferencích a pořádáním přednášek a workshopů [IBM, 2014].

### **Research at Google : Information Retrieval and the Web**

Je logické, že společnost Google, která je celosvětovou ikonou v oblasti webového vyhledávání, disponuje vlastním vývojovým oddělením v této oblasti. Výzkumu jsou podrobovány jak způsoby organizace a ukládání informací v prostředí webu, tak chování uživatelů při používání internetových vyhledávačů. Google vyvinul vlastní algoritmus pro řazení výsledků podle relevance, jehož podstatná část je neveřejná, aby se zamezilo manipulováním s výsledky vyhledávání. Členové oddělení se zároveň účastní vědeckých konferencí a publikují odborné články zabývající se vyhledáváním informací [GOOGLE, 2001a].

## Závěr

Vyhledávání informací prošlo jako obor i jako činnost během let značnými změnami. Od prvotních otázek, jakým způsobem využít k vyhledávání stroje, přes snahy zlepšit jejich výkon a přizpůsobit se uživatelským potřebám, až po myšlenku obohatit vyhledávače o vlastnosti typické pro člověka. Když si uvědomíme, kam až od svých počátků dospělo a co všechno má dnes běžný uživatel internetu k dispozici, musíme to nutně uzavřít tím, že se jedná o úžasný úspěch, přestože zdaleka nejsme na konci cesty.

K rozmachu a kontinuálnímu výzkumu v oblasti vyhledávání informací přispívá několik faktorů: jeho současná všeobecná dostupnost, potřeba šířit a sdílet lidské poznání a v neposlední řadě jeho komerční využitelnost. I kdyby výzkum nakonec skončil ve slepé uličce a ukázalo se, že některé inteligentní prvky ve vyhledávání, jako například hlasová konstrukce dotazů, se nestanou preferovanými, mohou alespoň posloužit určitým skupinám uživatelů. V případě hlasového ovládání vyhledávače například tělesně postiženým osobám se sníženou pohyblivostí rukou.

Pokud jde o dosažené mezníky, s úspěchem se povedlo do vyhledávání informací implementovat schopnost rozpoznávat objekty nebo tváře na obrázcích, rozpoznávat řeč a zpracovávat mluvený projev, automaticky zpracovávat přirozený jazyk, vyhledávat hudbu pomocí dotazů zkonstruovaných rovněž hudbou, nebo se učit. V žádné z těchto oblastí nebylo zatím dosaženo dokonalého stavu a všechny z nich jsou i nadále předmětem zájmu a výzkumu odborníků. Velký potenciál má uplatnění inteligentních prvků na mobilních telefonech a jiných přenosných zařízeních, kde může možnost komunikace s vyhledávačem pomocí hlasu nebo obrazu (např. fotografie pořízené přímo z přístroje) značně usnadnit a urychlit konstrukci dotazu.

Z některých prognóz formulovaných ve vlně nadšení z pokroků v oblasti umělé inteligence v 70. letech se už minimálně jedna vyplnila – počítače jsou dnes schopné v šachu porazit nejlepší lidské hráče [HUMAN-COMPUTER CHESS MATCHES, 2001]. Historie a nás učí, že mnoho z futuristických představ a teoretických vynálezů, které se objevily například v krásné literatuře, se jich nakonec opravdu realizovalo, za všechny jmenujme některé nápady Julese Verna, který byl skutečným vizionářem a ve svých dílech předpověděl například existenci raketového

pohonu nebo ponorky. Přechod od fantazie k praxi vždy vyžadoval čas, nicméně vyvodit se z toho dá minimálně to, že všechno je možné, dokud není jednoznačně prokázán opak. Proč by tedy nemohli v budoucnu existovat elektroničtí knihovníci či virtuální informační specialisté, kteří by v těch důležitých a výhodných aspektech kopírovali své lidské protějšky? Člověk je tvor čínorodý, zvědavý a ve svých snahách značně vytrvalý. Kdo ví, třeba si za nějakou dobu budeme moci se svým vyhledávačem popovídat.

Otázkou je, jestli inteligentní vyhledávání nenarazí na nějaké etické problémy. Při psaní práce jsem nenatrefila na žádný text nebo článek, který by se o této problematice zmiňoval, nicméně mně osobně to připadá docela důležité. Jestli budou vyhledávače komunikovat prostřednictvím umělých smyslů, získají od uživatelů vzorky jejich hlasu nebo obrazy tváří. Snadno tak budou moci monitorovat, kdo přesně daný rešeršní dotaz vytvořil. Mnohým uživatelům internetu dnes vadí i to, že jim vyhledávače personalizují vyhledávání na jejich počítači, a to přitom vychází z pouhého sledování navštívených stránek, které si vyhledávač spojuje s konkrétním přístrojem, nikoliv s konkrétní osobou.

Každopádně směr, kterým se vývoj v oblasti vyhledávání informací ubírá, je celkem jasně daný. Zapojení inteligentních prvků, schopnost vyhledávat netextové objekty podle obsahu nebo polidštění komunikace se systémem jsou momentálně výzvy, na které se vyhledávání informací snaží a pravděpodobně i dál bude snažit odpovědět.

## Seznam použité literatury

About. *Facesærch* [online]. [cit. 2014-04-21]. Dostupné z: <http://www.facesaerch.com/about.php>.

ACOUSTIC FINGERPRINT. 1981. In: *PCMag Encyclopedia* [online]. Computer Language Company, Inc, 1981-2014 [cit. 2014-04-14]. Dostupné z: <http://www.pcmag.com/encyclopedia/term/59679/acoustic-fingerprint>.

ANDREWS, Whit a Hanns KOEHLER-KRUENER. 2013. Magic Quadrant for Enterprise Search. In: *Gartner* [online]. Stamford, CT: Gartner, Inc. [cit. 2014-04-04]. Dostupné z: <http://www.gartner.com/technology/reprints.do?id=1-1FSGD38&ct=130528&st=sb>.

ARTIFICIAL INTELLIGENCE (AI). 2014. In: *Encyclopaedia Britannica* [online]. Encyclopaedia Britannica, Inc. [cit. 2014-03-18]. Dostupné z: <http://www.britannica.com/EBchecked/topic/37146/artificial-intelligence-AI>.

BAEZA-YATES, Ricardo a Berthier RIBEIRO-NETO. 1999. *Modern information retrieval*. Harlow: Addison-Wesley. ISBN 0-201-39829-X.

BALÍKOVÁ, Marie. 2003. Vyhledávání informací. In: *KTD: česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR [cit. 2013-07-03]. Dostupné z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000001618&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000001618&local_base=KTD).

BELEW, R. 1989. Adaptive Information Retrieval : Using a Connectionist Representation to Retrieve and Learn about Documents. In: *Proceedings of the Twelfth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval : Cambridge, Mass., USA, June 25-28, 1989*. New York: ACM Press. s. 11-20.

BERKA, Petr. 2003. Současné trendy umělé inteligence. *Acta Oeconomica Pragensia*. Vol. 11, no. 8, s. 41-46. Dostupné také z: <http://sorry.vse.cz/~berka/docs/4iz430/P00-TrendyAI.pdf>.

*BETH – ROZŠÍŘENÝ ONLINE KATALOG* [online]. 2008. Brno [cit. 2014-04-05]. Dostupné z: <http://knihomol.phil.muni.cz/katalog>.

BROOKS, H. M. 1987. Expert systems and intelligent information retrieval. *Information Processing*. Vol. 23, is. 4, s. 367-382. DOI: 10.1016/0306-4573(87)90023-9. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/0306457387900239>.

*CIIR – CENTER FOR INTELLIGENT INFORMATION RETRIEVAL* [online]. 2014. Amherst: University of Massachusetts [cit. 2014-04-04]. Dostupné z: <http://ciir.cs.umass.edu/>.

CROFT, W. Bruce. 1987. Approaches to intelligent information retrieval. *Information Processing*. Vol. 23, is. 4, s. 249-254. DOI: 10.1016/0306-4573(87)90016-1. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/0306457387900161>.



CROFT, W. Bruce, Donald METZLER a Trevor STROHMAN. 2010. Search engines: information retrieval in practice. Boston: Addison-Wesley. ISBN 0-13-607224-0.

CROFT, W. Bruce a Mark SANDERSON. 2012. The History of Information Retrieval Research. *Proceedings of the IEEE* [online]. 2012, Vol. 100, 100th anniversary issue, s. 1444-1451 [cit. 2013-07-03]. Dostupné z: <http://ciir-publications.cs.umass.edu/pub/web/getpdf.php?id=1066>. ISSN 0018-9219.

DIRAC : Detection and Identification of Rare Audio-visual Cues. 2013. *Carl von Ossietzky Universität Oldenburg* [online]. Oldenburg: Carl von Ossietzky Universität Oldenburg [cit. 2014-04-04]. Dostupné z: <http://www.dirac.uni-oldenburg.de/>.

EMOTIENT. 2014 About Emotient. *Emotient* [online]. San Diego: Emotient, Inc. [cit. 2014-04-02]. Dostupné z: <http://www.emotient.com/about>.

ENTERPRISE SEARCH. 2014. In: *Webopedia* [online]. QuinStreet, Inc. [cit. 2014-04-16]. [http://www.webopedia.com/TERM/E/enterprise\\_search.html](http://www.webopedia.com/TERM/E/enterprise_search.html).

*E-SPEAKING.COM* [online]. 2013. e-Speaking, LLC [cit. 2014-04-03]. Dostupné z: <http://www.e-speaking.com/>.

FACEBOOK. 2014. *Jakým způsobem Facebook navrhuje označení?* [online]. Facebook [cit. 2014-04-17]. Dostupné z: <https://www.facebook.com/help/122175507864081>.

FACIAL RECOGNITION SYSTEM. 2001. In: *Wikipedia : the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation [cit. 2014-04-04]. Dostupné z: [http://en.wikipedia.org/wiki/Facial\\_recognition\\_system](http://en.wikipedia.org/wiki/Facial_recognition_system).

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ VUT V BRNĚ. 2014. Oblasti výzkumu a výzkumné skupiny. *Fakulta informačních technologií VUT v Brně* [online]. Brno: Fakulta informačních technologií VUT v Brně [cit. 2014-04-04]. Dostupné z: <http://www.fit.vutbr.cz/research/groups/>.

FAKULTA INFORMATIKY MUNI. 1996. Umělá inteligence a zpracování přirozeného jazyka. *Masarykova Univerzita, Fakulta informatiky* [online]. Brno: Masarykova univerzita, 1996-2014 [cit. 2014-04-04]. Dostupné z: <https://www.muni.cz/fi/study/fields/8153?lang=cs>.

GOOGLE. 2001. *Google Search Appliance* [online]. Google [cit. 2014-03-24]. Dostupné z: <https://www.google.com/enterprise/search/products/gsa.html>.

GOOGLE. 2001a. *Information Retrieval and the Web* [online]. Google [cit. 2014-04-04]. Dostupné z: <http://research.google.com/pubs/InformationRetrievalandtheWeb.html>.

GOOGLE. 2001b. *Vše o vyhledávání* [online]. Google [cit. 2014-03-24]. Dostupné z: <http://www.google.com/insidesearch>.

GROSSMAN, David A. a Ophir FRIEDER. 2004. *Information retrieval : algorithms and heuristics*. 2nd ed. Dordrecht: Springer. ISBN 1-4020-3004-5.

GUPTA, Ayushi, Divya VERMA a Kajal GAMBHIR. 2012. Intelligent Information Retrieval. *Global Journal of Computer Science and Technology Software* [online]. Vol. 12, is. 12 [cit. 2013-07-14]. Dostupné z: [https://globaljournals.org/GJCST\\_Volume12/3-Intelligent-Information-Retrieval.pdf](https://globaljournals.org/GJCST_Volume12/3-Intelligent-Information-Retrieval.pdf). ISSN 0975-4172.

HABROVANSKÁ, Pavlína. 2010. Krátce o zpracování přirozeného jazyka. *Inflow: information journal* [online]. Roč. 3, č. 9 [cit. 2014-04-04]. Dostupné z: <http://www.inflow.cz/kratce-o-zpracovani-prirozeneho-jazyka>. ISSN 1802-9736.

HAJIČOVÁ, Eva. 1997. Zpracování přirozeného jazyka. In: MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ. *Umělá inteligence (2)*. 1. vyd. Praha: Academia. s. 244-256. ISBN 80-200-0504-8.

HLAVÁČ, Václav. 1997. Počítačové vidění. In: MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ. *Umělá inteligence (2)*. 1. vyd. Praha: Academia. s. 178-214. ISBN 80-200-0504-8.

HOFMANN, Thomas. 2006. *Machine learning in information retrieval* [online]. Darmstadt: Technische Universität Darmstadt. Dostupné z: <http://canberra06.mlss.cc/slides/Thomas-Hofmann.pdf>.

HUMAN-COMPUTER CHESS MATCHES. 2001. In: *Wikipedia : the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation [cit. 2014-03-04]. Dostupné z: [http://en.wikipedia.org/wiki/Human%E2%80%93computer\\_chess\\_matches](http://en.wikipedia.org/wiki/Human%E2%80%93computer_chess_matches)

CHOWDHURY, G. G. 2004. *Introduction to modern information retrieval*. 2nd ed. London: Facet. ISBN 1-85604-480-7. IBM. 2014. Information retrieval. *IBM* [online]. New York: IBM [cit. 2014-04-04]. Dostupné z: <http://www.research.ibm.com/haifa/dept/imt/ir.html>.

IEEE. 2014. About IEEE. *IEEE: advancing technology for humanity* [online]. IEEE [cit. 2014-04-04]. Dostupné z: <http://www.ieee.org/about/index.html>.

IEEE. 2014a. About IEEE Xplore® Digital Library. *IEEE Xplore® Digital Library* [online]. IEEE [cit. 2014-04-04]. Dostupné z: <http://ieeexplore.ieee.org/xpl/aboutUs.jsp>.

INSTITUTE OF FORMAL AND APPLIED LINGUISTICS. 2014. Grants. *Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic* [online]. Prague: Institute of Formal and Applied Linguistics [cit. 2014-04-05]. Dostupné z: <https://ufal.mff.cuni.cz/grants>.

JAIN, Anil K., Brendan KLARE a Unsang PARK. 2012. Face Matching and Retrieval in Forensics Applications. In: *IEEE Multimedia* [online]. s. 20 [cit. 2014-04-21]. ISSN 1070-986x. DOI: 10.1109/MMUL.2012.4. Dostupné z: [http://www.cse.msu.edu/~klarebre/docs/ieee\\_mm\\_forensic\\_fr\\_2012.pdf](http://www.cse.msu.edu/~klarebre/docs/ieee_mm_forensic_fr_2012.pdf). ISSN 1070-986X.

JANOVSKÝ, Dušan. 2012. Google PageRank : vzorec, vysvětlení a odpovědi. *Jak psát web: o tvorbě internetových stránek* [online]. [cit. 2013-07-14]. Dostupné z: <http://www.jakpsatweb.cz/seo/pagerank.html>.

JONÁK, Zdeněk. 2003. Komunikace člověk-počítač. In: *KTD: česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR [cit. 2013-07-03]. Dostupné z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000000477&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000000477&local_base=KTD).

JONES, Kevin P. (ed.). 1983. *Intelligent information retrieval : Informatics 7 : proceedings of a conference held by the Aslib Informatics Group and the Information Retrieval Group of the British Computer Society, Cambridge, 22-23 March 1983*. London: Aslib. ISBN 08-574-2187-5.

JORDAN, Michael I. a Stuart RUSSELL. 1999. In: *The MIT encyclopedia of the cognitive sciences*. Cambridge, Mass.: MIT Press, s. lxiii-xc. ISBN 0-262-73124-X.

KABINET INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ. 2004. Profil KISK. *KISK : Kabinet informačních studií a knihovnictví* [online]. Brno: Kabinet informačních studií a knihovnictví, 2004-2014 [cit. 2014-04-05]. Dostupné z: <http://kisk.phil.muni.cz/profil-kisk>.

KATEDRA INFORMAČNÍHO A ZNALOSTNÍHO INŽENÝRSTVÍ. 2000. Zaměření katedry. *Katedra informačního a znalostního inženýrství* [online]. Praha: Vysoká škola ekonomická v Praze, 2000-2014 [cit. 2014-04-04]. Dostupné z: <http://kizi.vse.cz/o-katedre/zanereni-katedry/>.

KATEDRA KYBERNETIKY ZČU. 2014. Vyhledávání informací. *Katedra kybernetiky ZČU* [online]. Plzeň: Katedra kybernetiky [cit. 2014-04-04]. Dostupné z: <http://ui.kky.zcu.cz/cs/research-fields/information-retrieval>.

KATEDRA POČÍTAČOVÝCH SYSTÉMŮ. 2013. *České vysoké učení technické v Praze, Fakulta informačních technologií* [online]. Praha: České vysoké učení technické v Praze Fakulta informačních technologií [cit. 2014-04-04]. Dostupné z: <https://www.fit.cvut.cz/fakulta/struktura/katedry/kps>.

KOGNITIVNÍ INFORMATIKA. 2000. *Vysoká škola ekonomická v Praze* [online]. Praha: Vysoká škola ekonomická v Praze, 2000-2014 [cit. 2014-04-04]. Dostupné z: <http://kogninfo.vse.cz/>.

KUBÁT, Miroslav. 1993. Strojové učení. In: MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ. *Umělá intelligence (1)*. 1. vyd. Praha: Academia. s. 168-183. ISBN 80-200-0496-3.

KUČEROVÁ, Helena. 2003. Umělá inteligence. In: *KTD: česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR [cit. 2013-07-03]. Dostupné z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000000137&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000000137&local_base=KTD).

LANGLEY, Pat a Herbert A. SIMON. 1995. Applications of machine learning and rule induction. *Communications of the ACM*. Vol. 38, is. 11, s. 55-64. Dostupné také z: <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA292607>.

LAPATA, Mirella a Frank KELLER. 2007. An Information Retrieval Approach to Sense Ranking. In: *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* [online]. Rochester [cit. 2014-04-04]. Dostupné z: <http://homepages.inf.ed.ac.uk/keller/papers/naacl07.pdf>.

LEVENE, M. 2010. *An introduction to search engines and web navigation*. 2nd ed. Hoboken, N.J.: John Wiley. ISBN 04-705-2684-X.

LOGIC THEORIST. 2001. In: *Wikipedia : the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation [cit. 2014-03-04]. Dostupné z: [http://en.wikipedia.org/wiki/Logic\\_Theorist](http://en.wikipedia.org/wiki/Logic_Theorist).

MACHINE VISION. 1999. In: *The MIT encyclopedia of the cognitive sciences*. Cambridge, Mass.: MIT Press, s. 501-503. ISBN 0-262-73124-X.

MAES, Pattie. 1994. Agents that reduce work and information overload. *Communications of the ACM*. Vol. 37, is. 7, s. 30-40. DOI: 10.1145/176789.176792. Dostupné také z: <http://dl.acm.org/citation.cfm?id=176792>.

MANCINI, John. 2010. *8 keys to findability* [online]. AIIM [cit. 2014-04-04]. Dostupné z [http://aiim.typepad.com/aiim\\_blog/2010/10/8-keys-to-findability.html](http://aiim.typepad.com/aiim_blog/2010/10/8-keys-to-findability.html).

MANNING, Christopher D., Prabhakar RAGHAVAN a Hinrich SCHÜTZE. 2008. *Introduction to information retrieval*. New York: Cambridge University Press. ISBN 978-0-521-86571-5.

MAŘÍK, Vladimír, et al. Úvod do problematiky umělé inteligence. *Umělá inteligence : automatizace*. 1988, roč. 31, č. 3, s. 2. MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ. 1993. *Umělá inteligence (I)*. 1. vyd. Praha: Academia. ISBN 80-200-0496-3.

MINKER, Wolfgang a Françoise NÉEL. 2002. Handling Knowledge Sources in Human-Machine Interaction. *International Journal of Speech Technology*. Vol. 5, is. 2, s. 171-188. DOI: 10.1023/A:1015476231853. Dostupné také z: <http://link.springer.com/article/10.1023%2FA%3A1015476231853>. ISSN 1381-2416.

MIZZARO, Stefano. 1996. *Intelligent interfaces for information retrieval : a review* [online]. Udine: University of Udine. Dostupné z: <http://sole.dimi.uniud.it/~stefano.mizzaro/research/papers/iirs.pdf>.

PAPÍK, Richard. 2001. Vyhledávání informací II. : uživatelské rozhraní a vlivy oboru "human-computer interaction". *Národní knihovna: Knihovnická revue*. Roč. 12, č. 1, s. 81-90. Dostupné z: <http://knihovna.nkp.cz/NKKR0102/0102081.html>. ISSN 1214-0678.

- PAPÍK, Richard. 2011. *Strategie vyhledávání informací a elektronické informační zdroje*. 1. vyd. Praha: Velryba. ISBN 978-80-85860-22-1.
- POTÁČEK, Jiří. 2003. Uživatelské rozhraní. In: *KTD: česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR [cit. 2013-07-03]. Dostupné z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000000070&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000000070&local_base=KTD).
- PROVOZNĚ EKONOMICKÁ FAKULTA. 2014. Inteligentní systémy a multimédia (GIM). *Provozně ekonomická fakulta Mendelovy univerzity v Brně* [online]. Brno: PEF MENDELU. Dostupné z: <https://ui.pefka.mendelu.cz/cs/oddeleni/GISYM>.
- PSUTKA, Josef. 2007. Hlasový dialog s počítačem. In: MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ. *Umělá inteligence (5)*. 1. vyd. Praha: Academia. s. 284-327. ISBN 978-80-200-1470-2.
- REGLI, T. 2008. Enterprise Search : Seek and Ye might Find. *Computers in Libraries*, Vol. 28, no. 7. s. 89-93. ISSN 10417915.
- ROBERTSON, S. E. a K. SPÄRCK JONES. 1976. Relevance weighting of search terms. *Journal of the American Society for Information Science*. Vol. 27, is. 3, s. 129-146. DOI: 10.1002/asi.4630270302. Dostupné také z: <http://www soi.city.ac.uk/~ser/papers/RSJ76.pdf>.
- SALTON, G., A. WONG a C. S. YANG. 1975. A Vector Space Model for Automatic Indexing. *Communications of the ACM*. Vol. 18, is. 11, s. 613-620. DOI: 10.1145/361219.361220. Dostupné také z: <http://portal.acm.org/citation.cfm?doid=361219.361220>.
- SIGIR. 2014. Fact sheet. *SIGIR : Special Interest Group on Information Retrieval* [online]. SIGIR [cit. 2014-04-04]. Dostupné z: <http://sigir.org/general-information/fact-sheet/>.
- SIRET RESEARCH GROUP* [online]. 2014. Prague: SIRET Research Group [cit. 2014-04-04]. Dostupné z: <http://siret.ms.mff.cuni.cz/>.
- SKLENÁK, Vilém. 2001. Data, informace, znalosti a Internet. Vyd. 1. Praha: C. H. Beck. ISBN 80-717-9409-0.
- SOUČKOVÁ, Martina. 2003. *Aspekty vztahu člověk-počítač s důrazem na uživatelské rozhraní*. Praha. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí práce Richard Papík.
- ŠVEJDA, Jan. 2003. Rešeršní systém. In: *KTD: česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR [cit. 2013-07-03]. Dostupné z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000001830&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000001830&local_base=KTD).
- ŠVEJDA, Jan. 2003a. Vyhledávání informací. In: *KTD: česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna

ČR [cit. 2013-07-03]. Dostupné z:

[http://aleph.nkp.cz/F/?func=direct&doc\\_number=000001825&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000001825&local_base=KTD).

TREE LEAF DATABASE MEW2010. 2010. *Department of image processing* [online]. Prague: Department of image processing [cit. 2014-04-04]. Dostupné z: [http://zoi.utia.cas.cz/tree\\_leaves](http://zoi.utia.cas.cz/tree_leaves).

ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ. 2014. Ukončené projekty. 2014. *Ústav informačních studií a knihovnictví, Filozofická fakulta Univerzity Karlovy v Praze* [online]. Praha: Ústav informačních studií a knihovnictví [cit. 2014-03-05]. Dostupné z: <http://uisk.ff.cuni.cz/listing.do?categoryId=14914>.

ÚSTAV INFORMATIKY AV ČR. 2013. O nás. *Ústav informatiky Akademie věd České republiky* [online]. Praha: Ústav informatiky AV ČR, v.v.i. [cit. 2014-04-04]. Dostupné z: [http://www.ustavinformatiky.cz/?id\\_jazyk=cs&id\\_stranky=o\\_nas](http://www.ustavinformatiky.cz/?id_jazyk=cs&id_stranky=o_nas).

VIDEO SEARCH ENGINE. 2001. In: *Wikipedia : the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation [cit. 2014-03-04]. Dostupné z: [http://en.wikipedia.org/wiki/Video\\_search\\_engine](http://en.wikipedia.org/wiki/Video_search_engine).

VOCŮ, O., 2012. Vyhledávání hudbou a jeho vazby na portály se sdíleným videem. *Knihovna* [online]. Roč. 23, č. 1, s. 63-83 [cit. 2014-04-04]. Dostupné z: [http://knihovna.nkp.cz/knihovna121/12\\_63.htm](http://knihovna.nkp.cz/knihovna121/12_63.htm). ISSN 1801-3252.

VOLF, Jaromír. 2007. Taktilní informace v systémech s umělou inteligencí. In: MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ. *Umělá inteligence (5)*. 1. vyd. Praha: Academia. s. 348-373. ISBN 978-80-200-1470-2.

*WIKIPEDIA : the free encyclopedia* [online]. 2001. San Francisco (CA): Wikimedia Foundation [cit. 2014-03-04]. Dostupné z: <http://en.wikipedia.org/>.

## Seznam obrázků

**Obr. č. 1** Schéma oborového začlenění inteligentního vyhledávání podle *Wikipedie*

**Obr. č. 2** Rozhraní vyhledávače *MUFIN* a část výsledků pro dotaz „tree“

**Obr. č. 3** Výsledky podobnostního vyhledávání ve službě *MUFIN* po výběru první fotografie na obr. č. 1 jako vzorové

**Obr. č. 4** Pracovní rozhraní nástroje *Search Google by Drawing*

**Obr. č. 5** Rešeršní dotaz a část výsledků vyhledávání v nástroji *Search Google by Drawing*

**Obr. č. 6** Zobrazení výsledků ve vyhledávači *Facesærch*

**Obr. č. 7** Rozhraní nástroje databáze *ChemsSpider* pro kreslení strukturních vzorců s nakreslenou molekulou

**Obr. č. 8** Výsledek vyhledávání v databázi *ChemsSpider* podle strukturního vzorce s přesnou shodou

**Obr. č. 9** 3D model styrenu v databázi *ChemsSpider*

**Obr. č. 10** Nástroj pro vyhledávání zahráním pomocí počítačové klávesnice v online hudební encyklopedii *Musipedia*

**Obr. č. 11** Nástroj pro vyhledávání zahráním pomocí flashové simulace piana v online hudební encyklopedii *Musipedia*

# **Příloha č. 1: Rešerše ve vybraných českých institucích na téma inteligentní vyhledávání**

*[v samostatném souboru]*