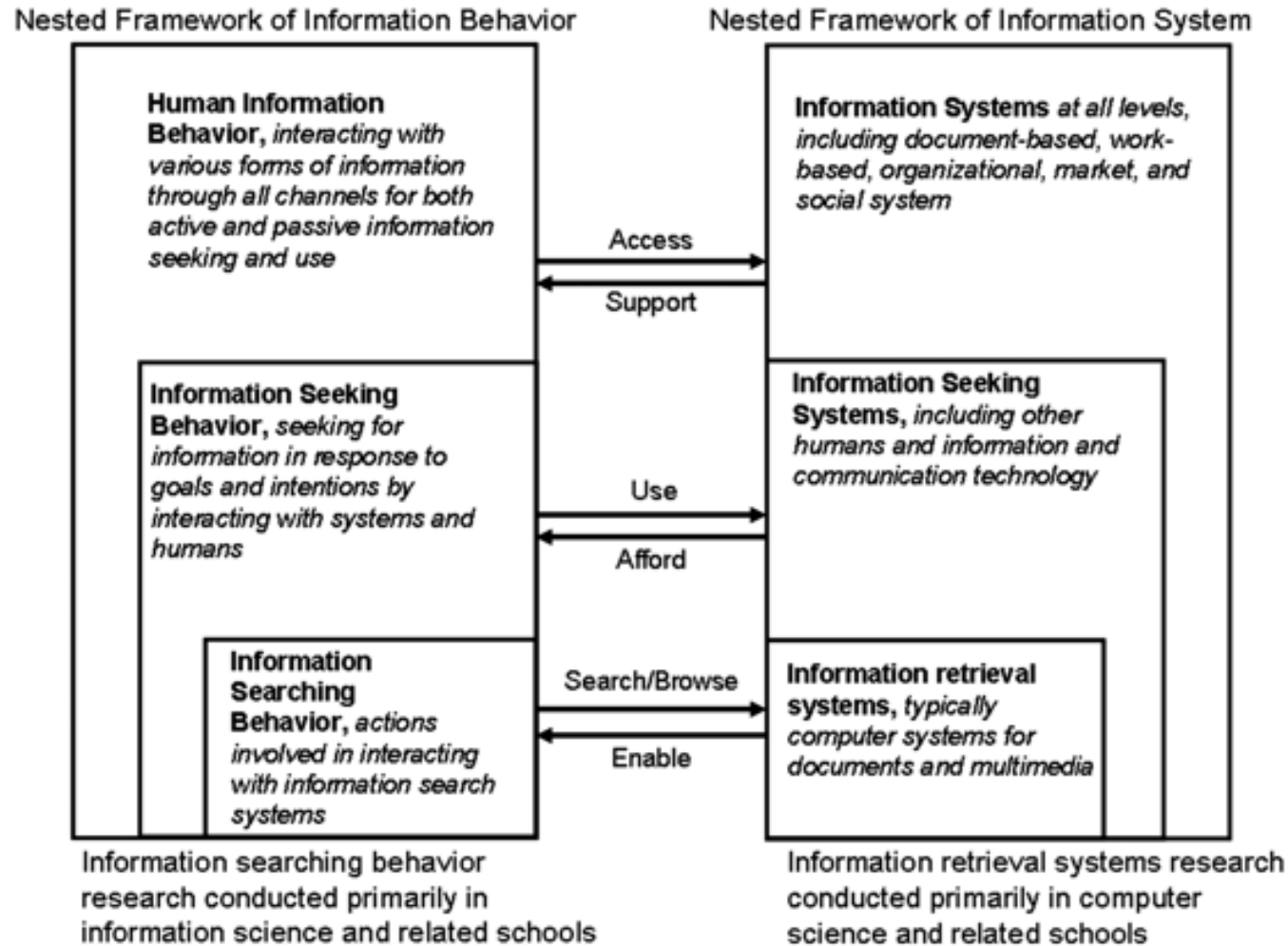


Vyhledávání informací

Vyhledávání informací

- interakce mezi lidmi a obsahem informačních systémů
- Information search – interakce lidí a vyhledávacího systému
- search - konkrétní chování lidí zabývajících se lokalizováním informací (vyhledávací strategie, posuzování relevance apod.)
- Information Retrieval – nalezení materiálů nestrukturované povahy z rozsáhlých sbírek uložených na počítači, které uspokojují informační potřeby
- retrieval – získávání informací z obsahu sbírky

Informační chování x vyhledávání informací



- obory informační chování a vyhledávání informací se vzájemně prolínají a ovlivňují
- poznatky z výzkumu informačního chování využitelné při designu systémů pro vyhledávání
- nové možnosti systémů vyhledávání informací mění informační chování a informační zvyky

Vznik oboru vyhledávání informací

- information retrieval – termín poprvé použil **Calvin Northrup Mooers** v roce 1950
- „informační chování řeší problém nasměrování uživatele k uložené informaci“
- spuštění družice Sputnik 1 Sovětským svazem (1957) – šok v U.S.A.:
 1. technologické zaostávání ve výzkumu vesmíru (spuštění programu Apollo)
 2. sputnik vysílá kódované signály, prolomení kódu trvá Američanům 6 měsíců, kódování však publikováno v sovětském časopise již dva roky před spuštěním družice a dostupné v anglickém překladu v amerických knihovnách

Vznik oboru vyhledávání informací

- **Weinbergova zpráva** (Alvin M. Weinberg, 1963) pro prezidenta Kennedyho o vědě, vládě a informacích – problém s informační explozí
- není potřeba jen vytvářet nové informace, ale také použít informace již existující (V. Bush, F. Kennedy) předpoklad: existence mechanismu pro vyhledání informace
- předpoklad: existence mechanismu pro vyhledání informace
- Kennedy volá po nové vědě zabývající se tříděním, referováním a syntetizováním informací
- American Documentation institut (ADI, zal. 1937) přejmenován na American Society for Information Science (ASIS, 1968)
- ve firmách je jednodušší, rychlejší a levnější udělat nový výzkum, než zjišťovat, co už bylo dříve uděláno

Průkopníci informačního vyhledávání

- **Hans Peter Luhn** – základ automatického indexování dokumentů na základě postupů textové statistiky
- 1950 – návrh technologií **SDI** (Selective Dissemination of information, česky ARI – Adresné rozšiřování informací, push služba) a **KWIC** (Keyword in Context – zvýraznění vyhledaných pojmů ve vyhledaném dokumentu)
- **Gerald Salton** – implementoval první vyhledávací systém, založený na počítání frekvence slov, vymyslel **model vektorového prostoru**, reprezentující dokumenty i dotazy jako vektory v n-dimenzionálním prostoru, reprezentujícího slova v dokumentu, vyhledáváno na základě podobnosti dokumentu a dotazu.
- V 60. letech se teorie stala východiskem pro konstrukci experimentálního systému vyhledávání **SMART** (System for the Mechanical Analysis and Retrieval of Text)

Průkopníci informačního vyhledávání

- **Eugene Garfield** – zvýšení vyhledatelnosti pomocí kopií obsahu časopisů (**Current Contents**). nápad reprezentovat cestu přenosu informací ve vědeckých časopisech pomocí **citačních indexů** (Science Citation Index).
- V roce 1960 založil ISI (Institute for Scientific Information) – první soukromý podnik provozující komerční systém vyhledávání informací
- **Norbert Henrichs** – se skupinou výzkumníků Siemens v Německu vytváří vyhledávací systém **GOLEM** (Großspeicherorientierte, listenorganisierte Ermittlungsmethode) - metoda vyhledávání založená na derivacích v hromadně uložených a do seznamu organizovaných dat, využití pro dokumentaci filozofické literatury).
- Systém **PASSAT** (Programm zur automatischen Selektion von Stichwörtern aus texten) - automatická selekce klíčových slov z textu na základě kmenů pojmů srovnávaných se slovníkem, používajícím váhy pojmů, nejvýše umístěné pojmy slouží jako preferované termíny)

Vznik informačního průmyslu

- **Roger K. Summit – DIALOG** (název označuje interaktivní systém mezi člověkem a strojem - 1966), vládní databáze dostupná veřejnosti, spuštění 1972, před nástupem PC, prohledávání přes telefonní terminál s vytáčeným připojením
- Summitův závěr – vyhledávání informací přes počítač může mít pozitivní efekt na výzkum
- **Carlos A. Cuadra – Orbit** (On-line Retrieval of Bibliographic Information Time-Shared), navazuje na systém CIRC vyvíjený od r. 1962, spuštění 1972, dnes Questel
- vyvíjí **SDC** (System Corporation Development)
- zaměřeno na databáze z přírodních věd a patenty (brzy přidána databáze z chemie - Chemical Abstract a fyziky – INSPEC)
- DIALOG komerčně zaměřený, Orbit zaměřený na výzkum a vývoj (byl proti DIALOGu efektivní a rychlý – ekonomický problém: účtován čas spojení, malý profit)

Vznik informačního průmyslu - medicína

- vývoj systémů pro vyhledávání informací původně pro letectvo, v 70. letech posun k medicínským informacím
- Národní knihovna medicíny tvoří lékařskou bibliografickou databázi MEDLARS
- od r. 1974 nabízena on-line
- **Medline** – vyhledávací systém pro databázi MEDLARS, spuštěno na základě **SDC** vyhledávacích služeb od r. 1972

Vznik informačního průmyslu - právo

- **OBAR** (Ohio Bar Automated Research) databáze právních dokumentů od r. 1965, vývoj Data Corporation, od r. 1968 Data Corporation, od r. 1970 OBAR činnosti outsourcovány Mead data Central (MDC)
- potřeba kvůli rozhodnutím soudů - precedenty
- vývoj fulltextového vyhledávacího systému - **Richard H. Giering**
- technologie: v textech odstraněny stop slova a vloženy do invertovaného souboru, Booleovské a proximitní operátory, Focus Command - zaměřený příkazový řádek (vyhledávání mezi nalezenými záznamy), základní informačně-lingvistická funkcionalita (pravidelné a nepravidelné formy plurálu), KWIC – Keywords in context (zde ještě Keywords in Color)
- 1973 – MDC spouští službu **LEXIS** (LEX Information Service) – vyhledávání právních textů
- Data Corp. vyvinula i vyhledávací služby pro archivy nakladatelství, např. Boston Globe – v r. 1977 nápad prodávat digitálně dostupné články z novin z databáze. Outsourcováno na MDC – služba **NEXIS** od r. 1980

Vyhledávání na WWW

- nástup PC a Internetu – období boomu ve vyhledávání informací, obnovený zájem o informační vědu
- vyhledávání všech druhů informací, ne jen vědeckých a odborných
- první systémy na vyhledávání: vyhledávače Yahoo! (1994), AltaVista (1995), Google (1998)
- systémy s vysokým výkonem pro instituce - Autonomy, Convera, FAST, Verity
- vývoj a výzkum vyhledávačů je řízený technologiemi, jádrem vývoj softwaru
- American Society for Information Science se opět přejmenovává na American Society for Information Science and Technology (ASIS&T)

Základní koncepty

• Informační potřeby (IP)

- **Konkrétní IP** (CIN – Concrete Information Need) – komunikace faktuálních informací, zvláštní případ navigace na WWW
- otázky: Jaké je hlavní město ČR? Jaká je URL UISKu?
- relevance informací uspokojujících CIN jasně definována
- **Na problém orientovaná IP** (POIN - Problem Oriented Information Need) – problém řešen na základě komunikace souboru dokumentů
- otázky: Jaký je vztah mezi managementem služeb a řízením kvality v obchodní administrativě? Jaké jsou různé interpretace Homunkula v druhé části Goethova Fausta?
- relevance se liší podle toho kdo a v jakém kontextu se

CIN

1. Thematic borders are clearly defined.
2. The search query can be expressed precisely.
3. Generally, one factual information is enough to satisfy the need.
4. Once the information has been transmitted, the information problem is resolved.

POIN

1. Thematic borders are *not* clearly definable.
2. The search query formulation allows for *several* terminological variants.
3. Generally, *various* kinds of documents must be acquired. Whether they will satisfy the information need is an open question.
4. The transmission of the information may modify the information problem or uncover a new need.

ptá

Základní koncepty

- **Informační potřeby**

- **objektivní IP** týká se objektivních faktických záležitostí, abstrahuje od konkrétního jedince, objektivně relevantní informace se může ukázat irelevantní pro konkrétního jedince (informaci již zná, zná jejího autora a nesouhlasí s ním, nemá čas číst tak dlouhý článek apod.)
- **subjektivní IP** – uvažujeme nad zvláštními podmínkami vyhledávaného předmětu
- **navigační IP** – cílem dotazu najít konkrétní webovou stránku
- **informační IP** – cílem dotazu najít konkrétní informaci ve statické formě, jediná očekávaná interakce je čtení, cílový dokument není tvořen jako odpověď na dotaz uživatele
- **transakční IP** – cílem dotazu najít stránku, kde dojde k další interakci – transakcím, které definují dotaz. Př. nakupování, zprostředkovatelské služby webu, stahování souborů, přístup do databáze apod.

Základní koncepty

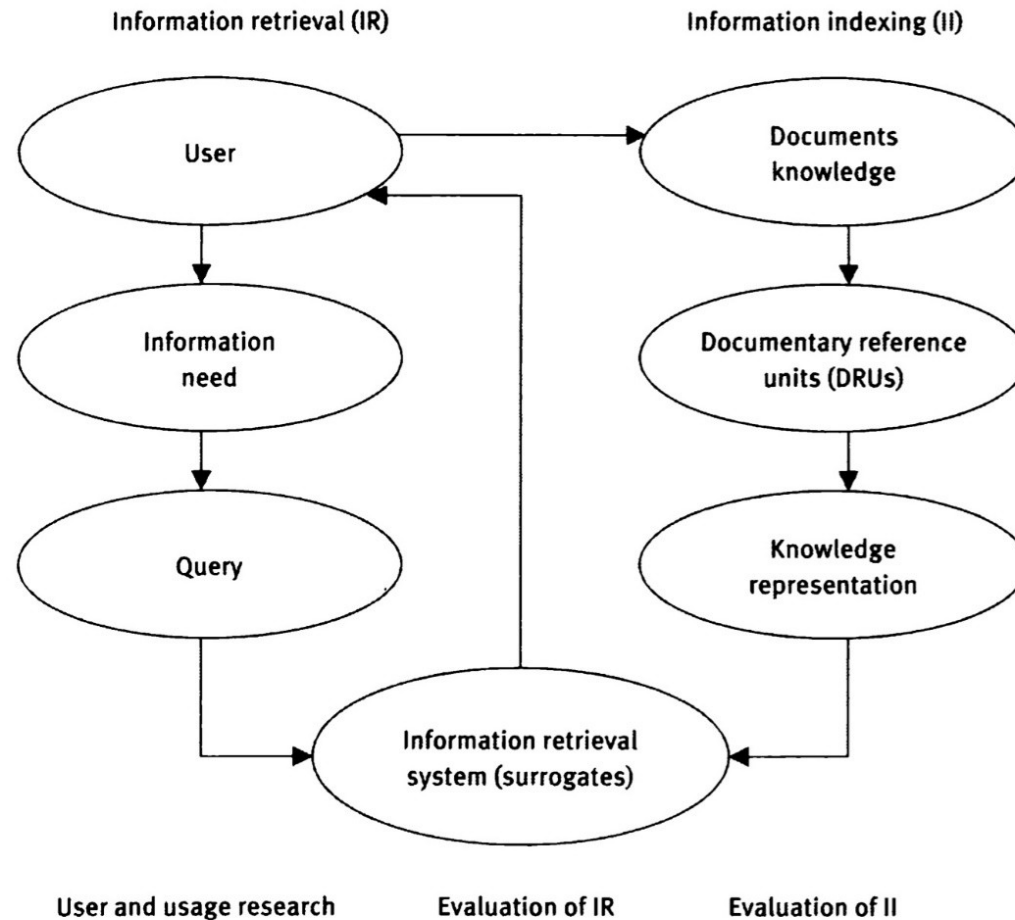
- **vyhledávací argument** – jeden či více výrazů a vyhledávacích parametrů, které jsou vyhledávány v dokumentu (u slov i jejich varianty, např. tvary množného čísla)
- **vyhledávací syntax** – vyhledávací systémy používají přibližně stejné množství příkazů, používají ovšem různou syntax
- Př. hledání informací o Julii Roberts ve filmu Notting Hill – nejprve musíme herečku a film znát (znalost dokumentů)
- dotazy:
 - HEADLINE: (“Julia Roberts” w/5 “Notting Hill”)
 - DIALOG: (Julia (n) Roberts AND Notting (w) Hill)/TI,
 - Google: “Notting Hill” “Julia Roberts” (přesně v tomto pořadí)

Základní koncepty

- **dokumentační referenční jednotka (DRU)** – reprezentace znalostí autora (kniha, článek, patentový dokument nebo interním záznam firmy apod.).
- **dokumentační jednotka (DU)** – slouží jako surogát, který reprezentuje obsah DRJ (obsahová redukce dokumentu, např. abstrakt, shrnutí) a jeho témata (např. klíčová slova). Tvoří informační specialista nebo systém pomocí automatického indexování, doplňují metadata. DU představuje základní přidanou hodnotu informace.
- **informační služby** – kombinace databáze (souhrn všech dokumentačních jednotek) a vyhledávacího systému (specializovaný softwarový program)

Indexování informací a vyhledávání informací

- hlavní problém vyhledávání – uživatel musí jasně formulovat a přesně to, co nezná
- 1. elementární znalost oblasti je základní požadavek (vědět o existenci hledaného fenoménu)
- 2. překlad informační potřeby do přirozeného jazyka
- 3. přeložit formulaci do dotazu, který bude umět použít vyhledávací systém – provede buď sám uživatel nebo to udělá automaticky systém za něj – zpracování přirozeného jazyka



- originální dokument segmentován do DRU. Výběr DRU do databáze dle vhodnosti. Kritéria katalogu: formální (př. jen anglické dokumenty, jen HTML), obsahová (jen dokumenty z chemie) Tvorba jednotky souvisí s cílovým rozsahem dat (př. celý film nebo epizody filmu).
- DRU je formálně popsána, obsahově transformována (př. abstrakt), objekt zpracován pomocí pojmů (thesaurus, nomenklatura, klasifikační systém). Prováděno intelektuálně (katalogizátor) nebo automaticky.
- Výsledkem DU (surogát), reprezentující DRU v informačních službách.

Základní koncepty

- **zprostředkování** - vyhledávací systémy specializovaných informačních služeb pracují na základě přesné shody (exact matching), uživatel proto musí znát příkazy syntaxe pro vyhledávání a vložit správný vyhledávací argument . Pokud účtován poplatek za každé kliknutí, stažení nebo čas připojení k databázi, vhodné využít informačního profesionála.

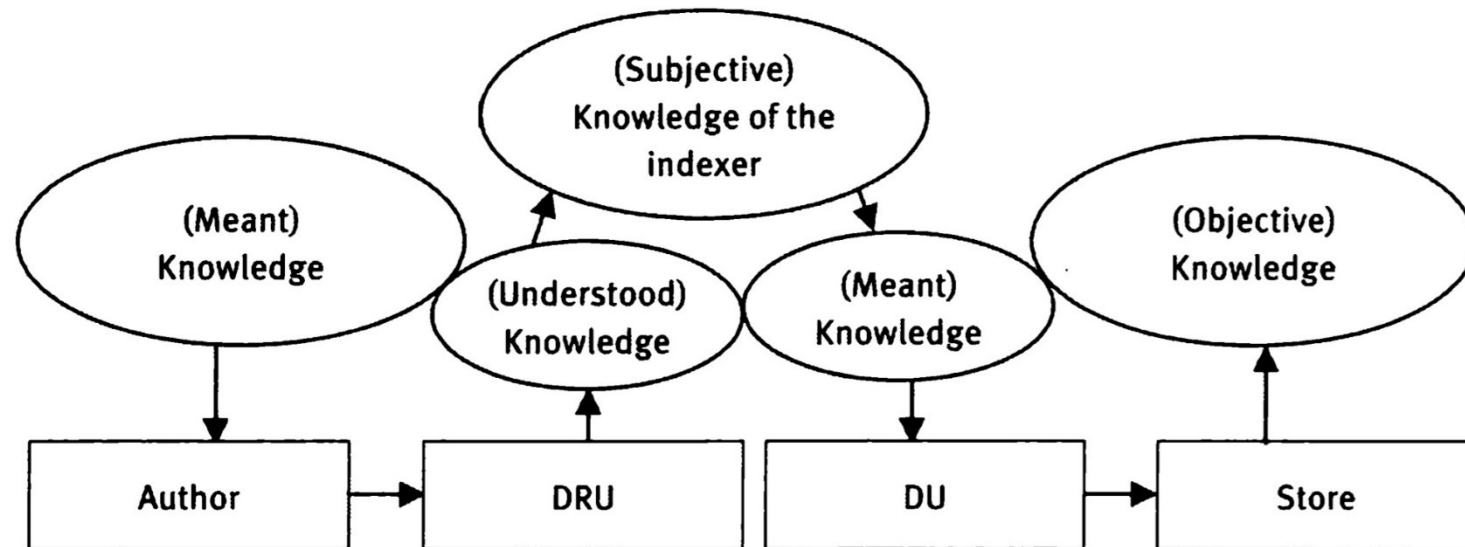


Figure A.2.7: Intermediation—Phase I: Information Indexing (here: with Human Indexer). DRU: Documentary Reference Unit, DU: Documentary Unit.

Display Settings: (v) Abstract

Send to: (v)

Am J Clin Pathol. 2009 Dec;132(6):824-8.

Pathology in the era of Web 2.0.

Schreiber WE, Giustini DM.

Department of Pathology & Laboratory Medicine, University of British Columbia, Vancouver, Canada.

Abstract

In the past few years, the term Web 2.0 has become a descriptor for the increased functionality of Web sites, including those with medical content. Most physicians do not know what Web 2.0 means or how it can impact their work lives. This review provides some background on the evolution of Web 2.0 and describes how its features are being incorporated into medical Web sites. Some potential applications of Web 2.0 in pathology and laboratory medicine are discussed, as are the issues that must be considered when adopting this new technology.

Comment in

Am J Clin Pathol. 2009 Dec;132(6):813-5.

PMID: 19926572 [PubMed - indexed for MEDLINE] Free Article

Publication Types, MeSH Terms

Publication Types

Review

MeSH Terms

Education, Continuing

Humans

Information Dissemination

Internet*

Laboratory Techniques and Procedures*

Medical Informatics/methods*

Pathology/education

Pathology/methods*

Search Engine

Telemedicine

Full Text FREE
Am J Clin Pathol

Related citations

Reflections on pathology and "Web 2.0".

[Am J Clin Pathol. 2009]

Pilot study of linking Web-based supplemental interpretive information I [Am J Clin Pathol. 2009]

Review HIV-associated resources on the internet. [Top HIV Med. 2009]

The medical school Web site: medical education's newest tool. [Isr Med Assoc J. 2000]

Review Informatics for practicing anatomical pathologists: marking a new [Mod Pathol. 2010]

See reviews...

See all...

Cited by 2 PubMed Central articles

Pathologists in a net-savvy world.

[J Pathol Inform. 2011]

Review Biomedical informatics and translational medicine. [J Transl Med. 2010]

All links from this record

Related Citations

Cited in PMC

Recent activity

Turn Off Clear

Pathology in the era of Web 2.0.

Základní koncepty

- Příklad dokumentační jednotky z databáze PubMed – záznam z lékařského časopisu
- bibliografická data (jméno časopisu, rok a měsíc zveřejnění, ročník, číslo, počet stran)
- v abstraktu kondenzována témata článku
- typy publikací – review
- PubMed používá intelektuální indexaci pomocí thesauru MeSH. Vypsané deskriptory slouží jako informační filtry, ty významné jsou označeny *

Figure B.2.2: Documentary Unit (Surrogate) of the Example of Figure B.2.1 in PubMed.

Služby push a pull

- jak se relevantní informace dostanou k uživateli? dva způsoby:
- uživatel aktivně získává informace ze systému (aktivní chování – pull služby, vyhledávání informací)
- uživatel pasivně čeká, až mu systém poskytne informace (pasivní chování – push služby, filtrování informací)
- formulace dotazů, které byly úspěšně zpracovány prostřednictvím služby pull, jsou uloženy do profilu
- oznámení a adresné rozšiřování informací automaticky (push služby) prohledávají databáze, a periodicky nebo v reálném čase upozorňují uživatele, když identifikují novou dokumentační jednotku dokumentu

Literatura

- Stock, W. G. a Stock, M. *Handbook of information science*. Berlin: De Gruyter Saur, 2015. ISBN 978-3-11-037364-6.