

Radka Římanová
State Technical Library
r.rimanova@stk.cz



Inspiration for the Presentation:

KEG Seminar April 23 2009

Jana Kravalová: Named entity recognition

viz http://keg.vse.cz/_slides/kravalova.pdf

Mention of difficulties with automatic recognition named

entities ...⇒

In the library (bibliographic catalogue) environment - solution to this problem \Rightarrow authority database

Could this tool help with an "institutional problem"?



Authority in Library

- Standardized (normative) form of keyword strings in the bibliographic database
- Long tradition the previous card catalogues, bibliographer's auxiliary apparatus
 - Core of Czech National Authority database the digitized card index of authors (it was built by bibliographer Jaroslav Kunc from the 30th of the 20th century)
- Library environment:
 - Anglo-American Cataloguing Rules2 (AACR2),
 - Machine Readable Cataloguing (MARC)
 - Z39.50 client-server protocol
 - Software catalogue modul of the library information system



Functional Requirements for Authority Data (FRAD), 2005-

publication of the final version of the study is expected

 Complements the Functional Requirements for Bibliograpic Records, IFLA, 1998.

Entity

person

family

corporate body

attributes + relations

name and/or identifier

controlled access points

User tasks - quality verification:

Find

Identify

Contextualize

Justify



"Nevertheless an authority file is not a biographical dictionary, nor a corporate bodies directory, etc., but where is the boundary?" Françoise Bourdon, 2001*

Czech answer:

Authority Database of the National Library of the Czech Republic 1996+

http://sigma.nkp.cz/F/?func=file&file_name=find-a&local_base=aut

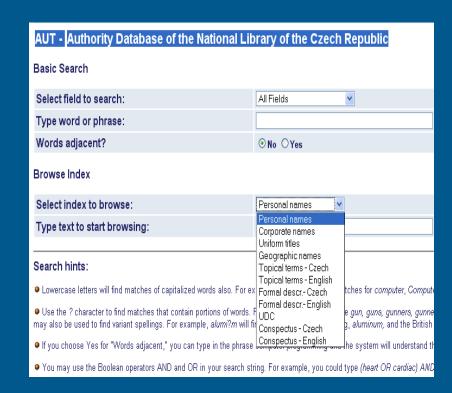
- From the beginning there is great emphasis on comments, notes, precise citing sources of the information, ...for above bibliographic standards
- Name authority record is not only the function name heading of a bibliographic record, but also the subject heading (term) for an access to the bibliographic record / full text

^{*}http://archive.ifla.org/IV/ifla67/papers/096-152ae.pdf



Authority Database of the National Library of the Czech Republic

- 1996+, 2000+ Cooperative project
- The database is created on the base of curently publish books or articals.
- Every day several dozen Czech librarians work to add and improve the database.





Portal "Národní autority ČR" http://autority.nkp.cz

- Methodology and coordination
 - Cooperative development and use of national authorities, project 2001+
 - National authorities in other memory institutions
 - Geolink.nkp.cz (link up records and maps via GPS) …



Authority Record of Institution Name Entity

AUT - Úplné zobrazení záznamu

Zvolte formát: Standardní -- MARC

----nz--a22----n--4500

Záznam 5 z 149

KA.

LDR

FMT

001	ko2006341891		
003	CZ-PrNK		
005	20060613082602.0		
800	060503 n acnnnaabnn-a a		
040	a ABA001 b cze		
1102	a Národní knihovna (Praha, Česko). Ib Ústřední ekonomick	ká knihovna 7 kd	2006341891
4102	a Ústřední ekonomická knihovna (Praha, Česko) 17 ko2006	6341891 O o	
4102	a Ústřední ekonomická knihovna NK 0 o		
4102	a Ústřední ekonomická knihovna Národní knihovny Ο ο	AUT - Úplné zob	razení zázna
4102	a ÚEK 7 ko2006341891 0 o	7 4 6 4 0	
4102	Ja ÚEK NK IO o	Zvolte formát: Standardní MARC Záznam 5 z 149	
5102	w a la Státní knihovna České socialistické republiky. Ib Ús	Ident. číslo	ko200634189
5102	w b ja Vysoká škola ekonomická v Praze. jb Centrum info	Záhlaví	●Národní kn
665	la Ústřední ekonomická knihovna byla součástí Národní (d	Odkaz. forma	Ústřední eko Ústřední eko
670	la www.(Centrum informačních a knihovnických služeb VŠE		Ústřední eko ÚFK
85642	u http://ciks.vse.cz/		ÚEKNK
906	a br20060613 b svsv	Předch. záhlaví Násled. záhlaví	Státní knihov
909	la CZ	Historie záhlaví	<u>Vysoká škola</u> Ústřední eko
	000341891	Zdroj informace	www(Centrun
SYS	000341091	Odkazy	http://ciks

AUT - Úplné zob	razení záznamu	
Zvolte formát: Star	ndardní MARC	
Záznam 5 z 149		
ldent. číslo	ko2006341891	
Záhlaví	 Národní knihovna (Praha, Česko). Ústřední ekonomická knihovna 	
Odkaz. forma	Ústřední ekonomická knihovna (Praha, Česko)	
	Ústřední ekonomická knihovna NK	
	Ústřední ekonomická knihovna Národní knihovny	
	ÚEK	
	ÚEK NK	
Předch. záhlaví	Státní knihovna České socialistické republiky. Ústřední ekonomická knihovna	
Násled, záhlaví	Vysoká škola ekonomická v Praze. Centrum informačních a knihovnických služeb	
Historie záhlaví	Ústřední ekonomická knihovna byla součástí Národní (dříve Státní) do konce r.1994. V r.1995 se stala základem nově vzniklého Centra informačních a knihov	
Zdroj informace	www(Centrum informačních a knihovnických služeb VŠE)	
Odkazy	http://ciks.vse.cz/	

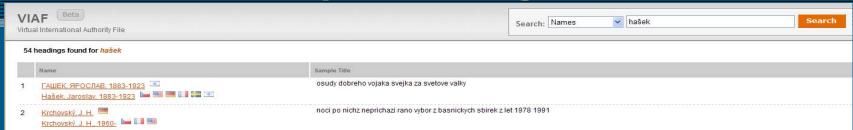
@ 2005 Ex Libris

Systém. číslo



Project Virtual International Authority File (VIAF)

Beta version - http://viaf.org/



- automated join of a few authority bases on the base of attributes or links to concrete bibliographic records
- Two level of attributes for right connection*:
 - precise properties: the work title, the standard number of works (ISBN, ISSN), exact biographical dates, contributions names, subject personal heading;
 - weaker properties: one of the biographical data (date of birth or death), the branch (can be expressed by word or code LCCS), the type of publications (books, films, music, ...), the language, the publisher, the works sub-title, the date of publication, the country of operation, the relator role

^{*} List of specific fields in enhanced authority record VIAF - VIAF Tags viz http://viaf.org/html/viaftags.html
The Knowledge Engineering Group Seminar, June 11 2009



Czech Authority Database

- Authority records are risen, corrected, added, deleted and combined during the processing (cataloging) of print and electronic library holdings and thanks to it they contain important and verified data about a concrete entity.
- Library standards are tradicional rules, changes are slow but they exist and they are required by the creators of these databases.
- Cooperative project is funded from public budget, respects heterogeneous software environment. Its strong value-added is high level of knowledge of Czech cataloguers in the area of building factual databases.



Authority files could be used in other informatics projects.

Authority files derive benefit from the other use.