

Dlouhodobá ochrana digitální dokumentů

Mgr. Jaroslav Kvasnica

Obsah přednášky

1. Digitální dokument a metadata
2. Dlouhodobá ochrana digitálních dokumentů
3. Webová archivace

Digitální dokument

Digitální dokument je dokument kódovaný v binární soustavě a přesně interpretovaný pouze pomocí výpočetní techniky.

Vlastnosti digitálního dokumentu

- **kopie = originál**
- nosič a informace nejsou na sobě závislé
- používáním se neopotřebovává
- informace není přístupná přímo
- instantní zničení

Vznik digitálních dokumentů

- digitalizace (konverze z analogové formy)
- born-digital
 - statické x dynamické

Metadata

data o datech

popis (čehokoli)



▶ ⏩ 🔊 1:20 / 2:23

CC ⚙️ HD 📺 🗉

French Bulldog Puppy Tricks 12 wks old Brody Brixton



MeSoMiniMisa

▶ Subscribe 71,552

6,117,223

+ Add to ➦ Share ⋮ More

👍 20,045 🗨️ 674

Published on Jan 1, 2015

What can a puppy learn in 4 days? :) This Frenchie pup will melt your heart.

Misa Minnie's new little bro Brody Brixton will show you what a few days of training with positive



data



title



French Bulldog Puppy Tricks 12 wks old Brody Brixton

author

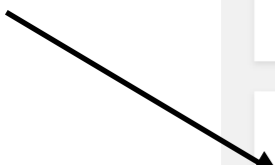


MeSoMiniMisa

Subscribe 71,552

6,117,223

date

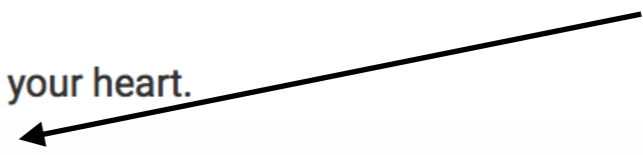


Published on Jan 1, 2015

What can a puppy learn in 4 days? :) This Frenchie pup will melt your heart.

Misa Minnie's new little bro Brody Brixton will show you what a few days of training with positive

description

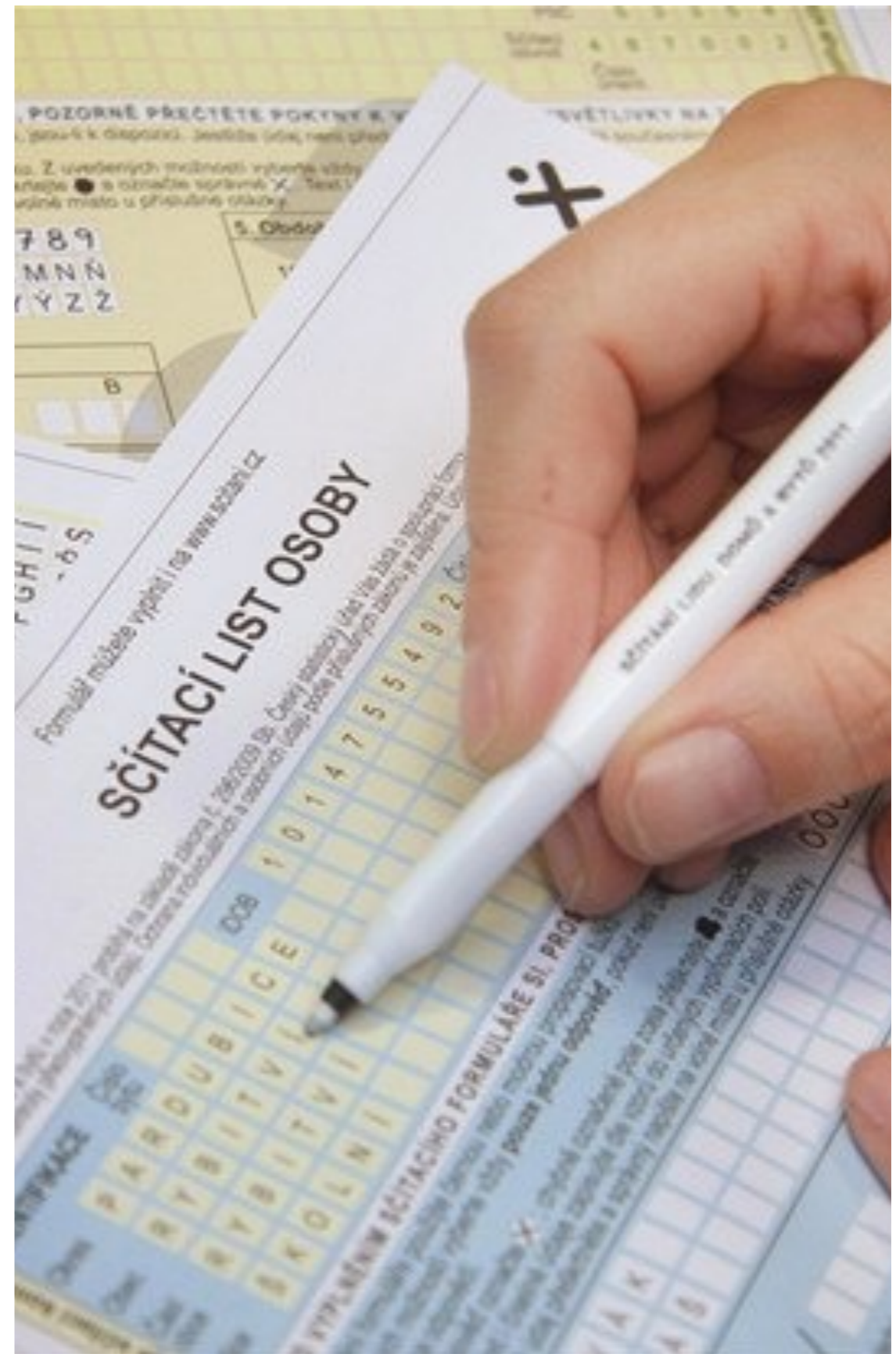


+ Add to ➦ Share ⋮ More

👍 20,045 💬 674






*Data nebo
metadata???*



Základní výsledky

 Tabulka
  Graf
  Mapa

 XLS
  XML
  Tisk

Vyberte území Česká republika

Období: 26.3.2011

ČR, kraj	Obyvatelstvo celkem	v tom		z toho podle věku			Ekonomicky aktivní	z toho zaměstnaní	Obydlené domy	Obydlené byty	Hospodařící domácnosti
		muži	ženy	0 - 14	15 - 64	65 a více					
Česká republika	10 436 560	5 109 766	5 326 794	1 488 928	7 267 169	1 644 836	5 080 573	4 580 714	1 800 075	4 104 635	4 375 122
Hlavní město Praha	1 268 796	613 738	655 058	153 622	908 321	201 029	644 643	600 730	92 927	542 168	579 509
Středočeský kraj	1 289 211	637 252	651 959	199 300	895 024	190 911	639 851	587 539	286 780	482 860	523 045
Jihočeský kraj	628 336	308 296	320 040	91 119	435 187	100 000	307 130	280 844	123 048	247 608	262 692
Plzeňský kraj	570 401	282 137	288 264	79 469	396 468	92 734	278 674	255 278	105 835	226 298	242 397
Karlovarský kraj	295 595	145 483	150 112	42 159	207 480	44 538	139 871	123 100	39 845	119 403	128 904
Ústecký kraj	808 961	397 453	411 508	121 692	565 358	117 899	377 298	325 987	115 679	330 981	352 346
Liberecký kraj	432 439	211 537	220 902	64 597	301 267	64 949	208 512	185 944	73 380	171 328	183 299
Královéhradecký kraj	547 916	268 967	278 949	79 127	374 898	92 020	260 819	238 986	109 736	215 277	228 256
Pardubický kraj	511 627	252 310	259 317	75 093	352 543	82 330	246 779	224 590	104 850	196 288	207 396
Kraj Vysočina	505 565	250 196	255 369	73 798	347 663	82 741	243 720	220 743	108 062	188 191	198 504
Jihomoravský kraj	1 163 508	567 882	595 626	162 794	808 861	188 684	572 019	511 844	225 006	443 358	473 520
Olomoucký kraj	628 427	305 526	322 901	90 398	434 573	101 647	303 992	268 013	118 882	243 624	257 964
Zlínský kraj	579 944	282 500	297 444	82 267	399 942	96 398	281 576	252 815	120 444	217 093	229 682
Moravskoslezský kraj	1 205 834	586 489	619 345	173 493	839 584	188 956	575 689	504 301	175 601	480 158	507 608

Sčítání lidu, domů a bytů 2011 - definitivní výsledky

e-04000-12	Obyvatelstvo podle Sčítání lidu, domů a bytů 2011 - Česká republika	
e-05000-12	Základní výsledky Sčítání lidu, domů a bytů 2011 - Česká republika	
e-07000-12	Obyvatelstvo podle Sčítání lidu, domů a bytů 2011 - Česká republika a kraje	
e-08000-12	Základní výsledky Sčítání lidu, domů a bytů 2011 - Česká republika a kraje	
e-09000-13	Domácnosti podle Sčítání lidu, domů a bytů 2011 - Česká republika	
e-10000-13	Domácnosti podle Sčítání lidu, domů a bytů 2011 - Česká republika a kraje	
e-12000-13	Základní informace o vybraných územních celcích podle SLDB 2011 - ČR, kraje, okresy, správní obvody ORP a obce (včetně městských částí územně členěných statutárních měst)	
e-13000-13	Základní informace o krajích a velikostních skupinách obcí podle výsledků Sčítání lidu, domů a bytů 2011 - Česká republika	
e-14000-13	Byty podle Sčítání lidu, domů a bytů 2011 - Česká republika	

Typy metadat

- popisné (co je to za objekt, jak jej nalézt)
- administrativní (jak objekt spravovat)
- technické (technický popis objektu)
- strukturální (vnitřní uspořádání objektu)

MARC

<http://marc-must-die.info/>

XML!

- Dublin Core, MODS, PREMIS, MIX, METS a jiné...

MARC21

245 \$b

600

534 \$c, 760 762, 765, 767, 770, 772, 773, 774, 775, 776, 777, 780, 785, 786, 787 \$d

020 \$a

MODS

<subTitle>

<name>

<publisher>

<identifier type="isbn">

Dlouhodobá digitálních dokumentů

Digital preservation is a formal endeavor to ensure that digital information of continuing value remains accessible and usable.



Our children will **never** know
the link between the two

aprilroad.com

Cíle dlouhodobé ochrany digitálních dokumentů

uchování; zpřístupnitelnost; srozumitelnost

Technická rizika

- degradace nosiče
- zastarávání nosiče
- poškození (přírodní katastrofy, viry, zálohovací procesy)

Logická rizika

- zejména formátová rizika
 - zastarávání souborových formátů
 - vlastnictví formátu
 - specifikace formátu
- dokumentace
 - srozumitelnost, rozpoznatelnost

Systemová rizika

- financování!
- integrita -> dostanu nezměněný dokument? Falšování, poškození, cenzura
- autenticita -> dostanu opravdu dokument, který jsem žádal? Má takto dokument vypadat?
- **důvěryhodnost repozitáře**

Technologická dlouhodobá ochrana

- výběr vhodného nosiče
- bitová ochrana
- zálohování

Logická dlouhodobá ochrana

- Enkapsulace
- Migrace
- Emulace
- Technické muzeum

Migrace

- výměna zastarávajícího formátu za nový
- výběr formátu: signifikantní vlastnosti dokumentu
- dokumentace formátu

+ uživatelé pracují se svým systémem, široká aplikace

- chybovost, nutnost pravidelného opakování, fin. náročnost

Emulace

- simulace hardwarového nebo softwarového prostředí na jiném systému
-

+ uživatel pracuje v původním systému, nezasahuje se do sekvence bitů

- uživatel pracuje v původním systému , uživatel musí používat speciální sw, nové platformy = nové emulátory,

<http://oldweb.today>

Webová archivace

Web archiving is the process of collecting portions of the World Wide Web to ensure the information is preserved in an archive for future researchers, historians, and the public.

Proč archivovat web?

“Je snazší nalézt exemplář filmu z roku 1924, než webové stránky z roku 1994.”

M.S. Ankerson. “Writing web histories with an eye on the analog past.” 2012.

<http://nms.sagepub.com/content/14/3/384.full.pdf+html>

Proč archivovat web?

- proč archivovat zrovna web?
 - ochrana netištěné kulturní, umělecké a historické produkce pro další generace
 - přesun publikační činnosti na internet
 - prchavost elektronických zdrojů

Typy webu

- **Surface web** - běžně přístupná část webu
- **Deep web** (invisible, hidden) - dynamický obsah, kontextuální, privátní, skriptovaný atd.
- **Dark web** (darknet) - obvykle nelegální obsah, těžko dosažitelný (P2P)

Typy archivace

Personální

- pro osobní potřebu
- firemní archivace
- neřeší rozdíly mezi surface x deep

Institucionální

- za účelem uchování kulturního dědictví
- knihovny, univerzity
- často web-scale archiving

Web scale archiving

- webové archivy se snaží archivovat tak velkou část internetu, že není možné kontrolovat akvizici, ochranu, zpřístupnění ani kontrolu kvality jednotlivých webových stránek pouze lidskými silami.
- Internet Archive, národní knihovny, univerzity...

Webarchiv

- www.webarchiv.cz
- digitální archiv **českých** webových zdrojů
- český web = “všechny” bohemikální zdroje

Webarchiv

- stáří archivu: 3. 9. 2001
- > 250 TB
- webarchiv.cz
- přes 1 300 000 domén .cz

Zpřístupnění

- Wayback Machine (zobrazování)
- Vyhledávání:
 - primárně WWW
 - katalog Aleph
 - Wayback Machine

- *méně než 1 % zdrojů zpřístupněných online*
- celý archiv dostupný pouze v Klementinu
- Bude to lepší v budoucnu? **Ne.**
 - terminály v dalších knihovnách?
 - **metadata**

Děkuji za pozornost!

Jaroslav Kvasnica
jaroslav.kvasnica@nkp.cz

