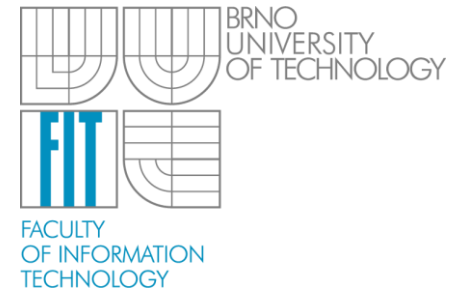


Využití ontologií k popisu vizuálních rysů dokumentu

Martin Milička a Radek Burget

Brno University of Technology, Faculty of Information Technology
Božetěchova 2, 612 00 Brno, CZ

imilicka@fit.vutbr.cz
burgetr@fit.vutbr.cz



WIKT 2012, Smolenice

HTML dokument

- stromový popis dokumentu (+ definice vizuálního vzhledu)
- x stromový popis obvykle neodpovídá vizuálnímu vzhledu

Vizuální informace obsahuje *více logických vazeb* než jsme schopni ve stromové struktuře zaznamenat.

Zajímavé informace:

- sousednost, logické vazby (nad, pod, atd.)
- zapouzdření
- atd.

Řešení:

- reprezentace obecným grafem
- implementace RDF + Ontologie

iDNES.cz / Zprávy

Pondělí, 5. listopadu 2012, Svátek má Miriam.


iDNES.cz Zprávy Kraje Sport Kultura Ekonomika Finance Bydlení Cestování Auto Hobby Mobil Technet Ona Xman Revue Blog Video Hry vice

Volby 2012 Domáci Zahraníčí Černá kronika Důma Čtenářů Kávárna Kemej MF DNES Počasí reklama Firemní články

Týden bude deštivý a chladný, hřebeny hor posype sníh **A**

5. listopadu 2012 7:14

Zádný slunečný podzim neslibují meteorologové pro následující týden. Naopak má být hlavně oblačno a zataženo s tím **B**mi převážně do deseti stupňů. Na hřebenech hor se objeví sníh. Nijak zvlášť se počasí nelepší ani o víkend.



Pres týden nás čeká oblačno, přeháňky a déšť. (ilustrační snímek) | foto: Alexandr Satinský, MF DNES

Další články z Oubřiky

Outsider prezidentské volby loží podpisy na veřejných **D**hodcích

Nová senátorka Milada Emmerová z ČSSD odpovídá čtenáři **E**iDNES.cz

Příběhy 20. století: Komunistům nikdy nevěřil, skončil **F**ranových dotech

OBRAZEM: Čtenáři rozehnali inverzi, podívejte se **G**chvatně snímky

P

Další články z rubriky

Už bylo dost aplyých doktorů, našla se komora. Řeší, co s tím

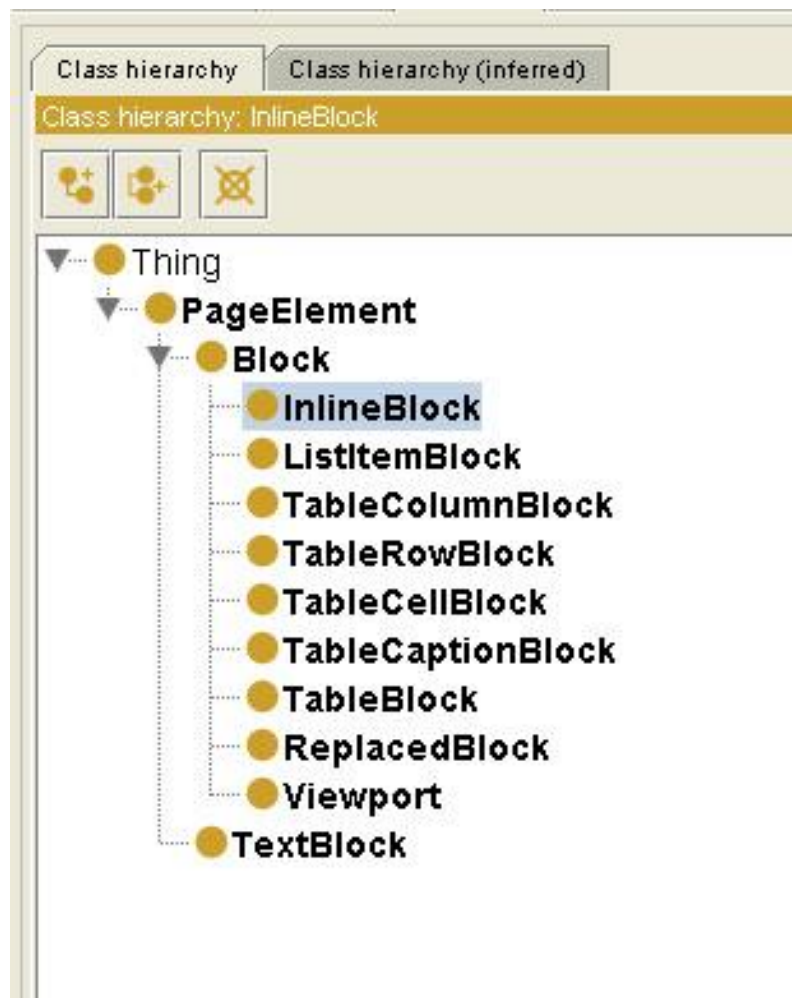
OBRAZEM: Viažný rozejzd a pak sedup. Jaký vlastně byl kongres ODS?

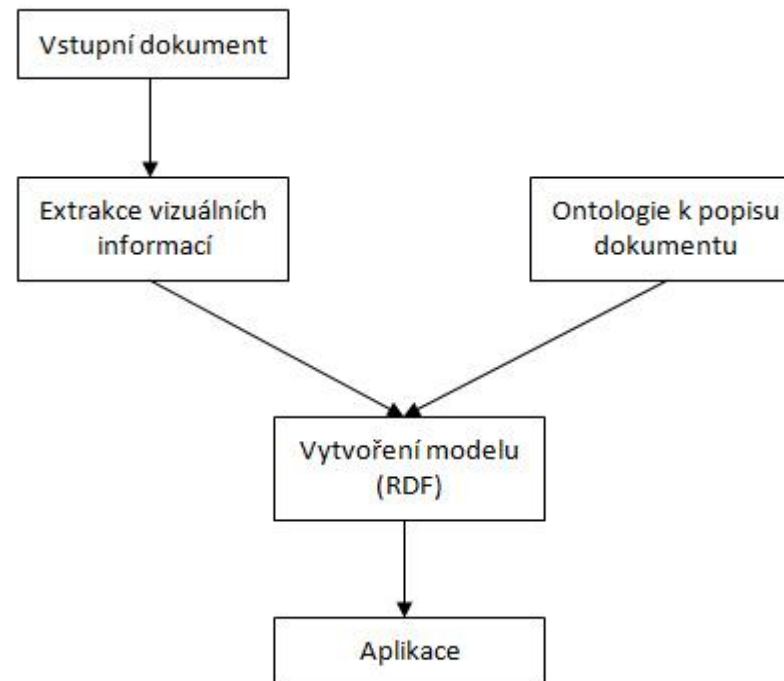
Třídy:

- PageElement
- Block
- TextBlock
- ReplacedBlock
- atd.

Specifické vlastnosti tříd:

- P parentOf C
- C isNeighbor D
- D isBelow C
- C isAbove D
- A leftOf C
- C rightOf A
- atd.





Extrakce vizuální informace se provádí z vyrenderovaného dokumentu. Kromě jiného je tak možné zjistit vzájemné pozice jednotlivých elementů.

- Extrakce informace z layoutu, atd.
- Porovnávání dokumentů na základě vizuálních rysů
- Vazba na ontologie
 - odvozování a hledání vzorů (existující reasonery)
 - získávání pokročilých znalostí
 - doplnění znalostí – anotace
 - na základě získaných znalostí - propojení s jinou ontologií (ontologie osob)

- Popis vizuální informace dokumentu pomocí RDF a ontologie
- Reprezentace vzájemných vazeb mezi elementy stránky (parentOf, leftOf, upOf, hasColor, atd.)
- Přesnější informace o jednotlivých informacích
- Možnost využití existujících reasonerů k odvozování
- Doplnění znalostí na základě definované anotace
- Možnost propojení s jinou ontologií (ontologie osob)

Děkuji za pozornost!