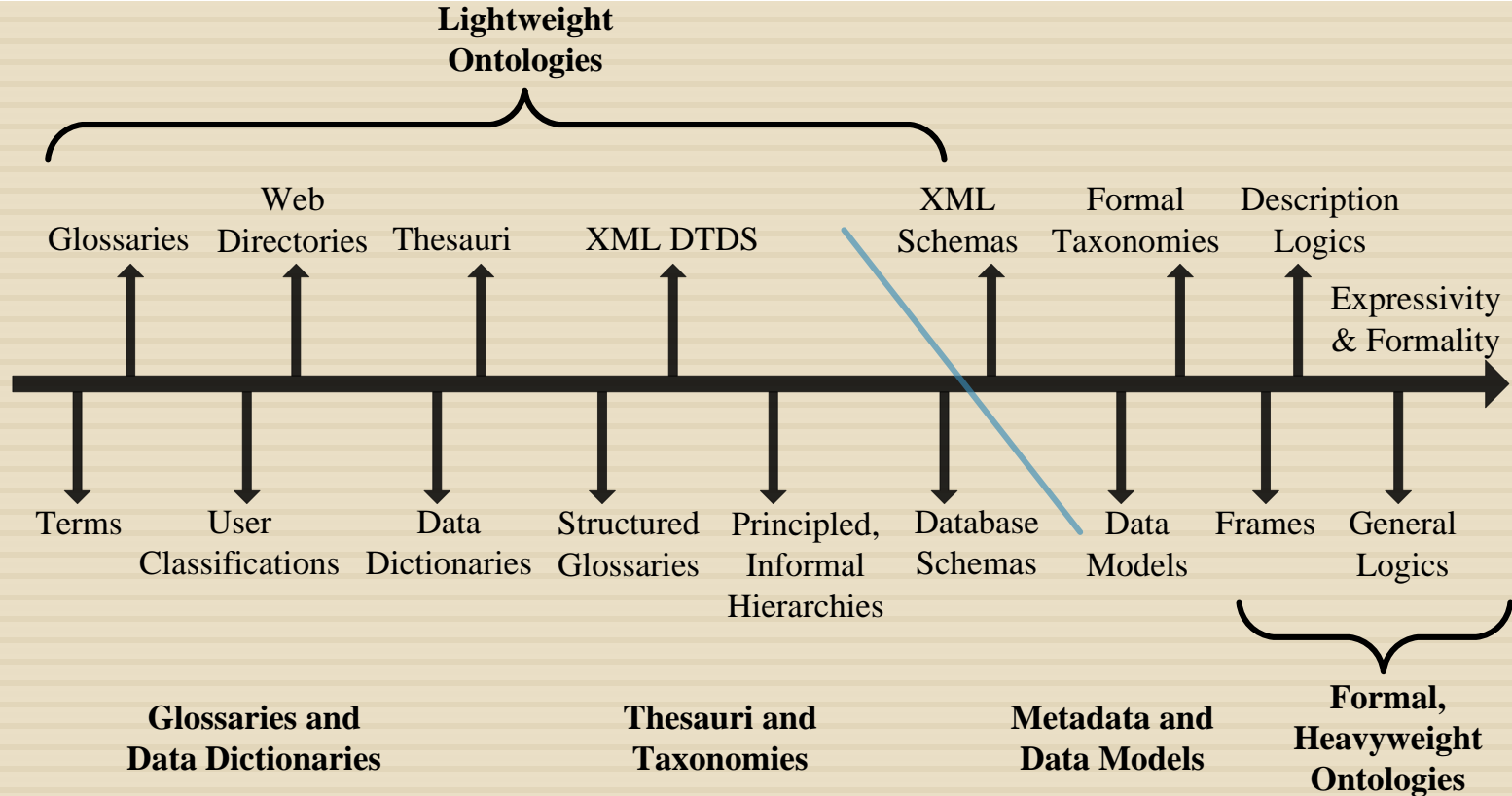


ÚDRŽBA ĽAHKEJ SÉMANTIKY
REPREZENTOVANEJ
INFORMAČNÝMI ZNAČKAMI

Metadáta informačných priestorov

- Pridávajú dodatočné informácie k obsahu
- Typy metadát (NISO 2004)
 - Opisné metadáta
 - Štruktúralne metadáta
 - Administratívne metadáta
- Často reprezentujú sémantiku

Formy ontol\u00f3gi\u00ed (Wong et al. 2011)



Ľahké ontológie (Giunchiglia & Zaihrayeu 2007)

- Reprezentujú základné znalosti
- Založené len na
 - ▣ *Konceptoch* – napr. kľúčových slovách
 - ▣ *Jednoduchých reláciách* – napr. is-a, part-of
- Napr. taxonómie, slovníky
- Postačujú pre väčšinu úloh informačných systémov

Projekt PerConlK

5

- Ciel': Podpora vývoja
- Informačný priestor softvérovej firmy
- Zbierané metadáta
 - Aktivity vývojárov (frekvencia zmien zdrojového kódu, študované dokumentácie, prezerané fóra, ...)
 - Informácie založené na obsahu informačného priestoru (vzory, pachy, počet riadkov, ...)

Otvorené problémy

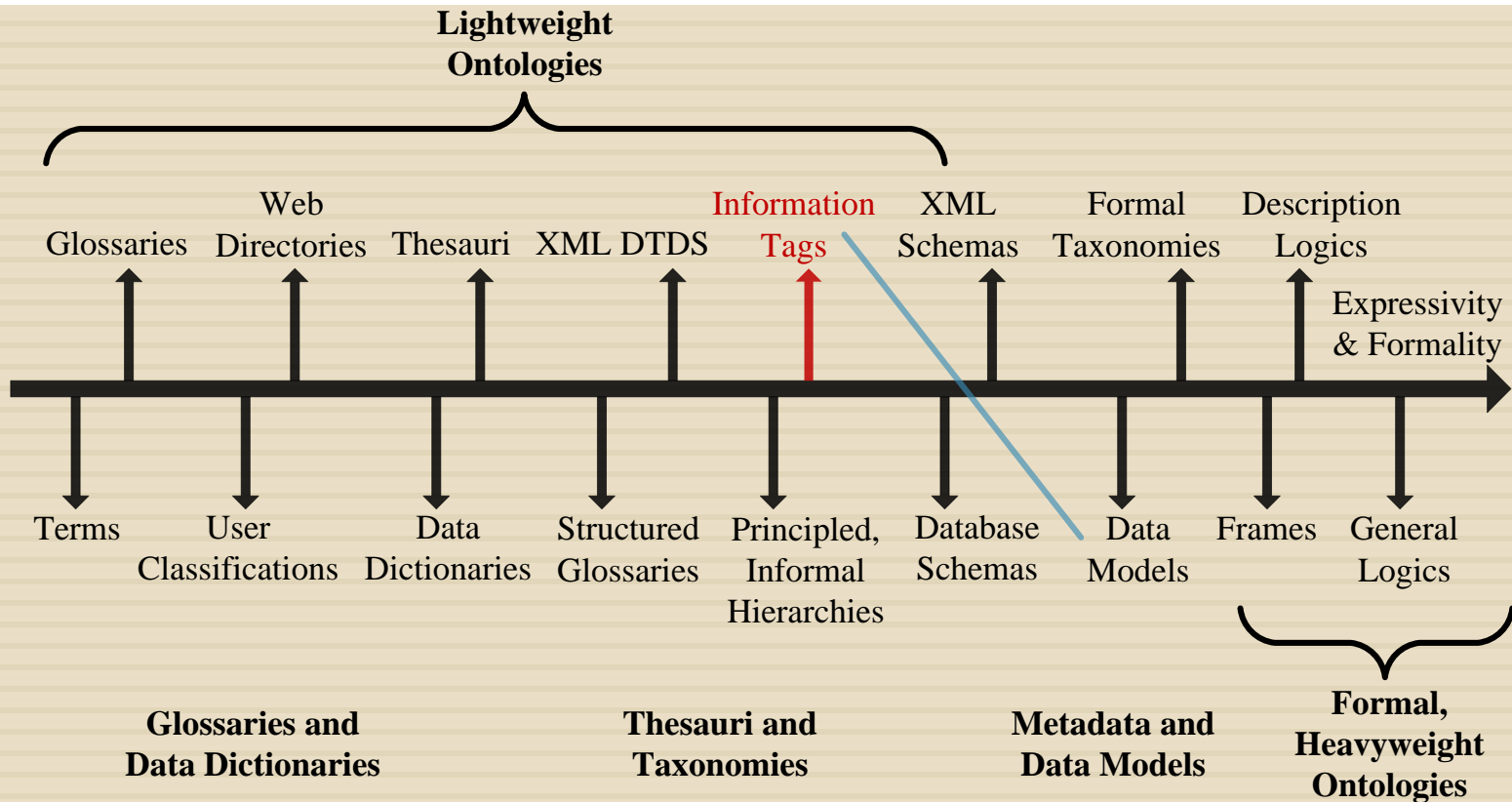
- Ako udržiavať metadáta?
 - ▣ Zdrojový obsah nie je statický
 - ▣ Metadáta sú závislé od zdrojového obsahu a času
- Ako modelovať metadáta?
 - ▣ Efektívna údržba
 - ▣ Možnosť dopytovania

Informačné značky

- Koncept značiek/anotácií používateľov
- Opisné metadáta s definovaným vzťahom voči označenému informačnému artefaktu
- Trojice:
 - *Typ* – význam informačnej značky
 - *Ukotvenie* – označený informačný artefakt
 - *Telo* – štruktúrovanú informáciu

Informačné značky ako ľahká ontológia

8



Problémy údržby metadát

- Platnosť ukotvenia
 - ▣ Riešiteľné vhodným robustným deskriptorom (Phelps & Wilensky 2000)
- Platnosť nesenej informácie
 - ▣ Zložitý problém, potreba porozumieť štruktúre a významu metadát
 - ▣ Metadáta viazané len na určitú verziu dokumentu (Sanderson & Van de Sompel 2010)

Údržba informačných značiek

- Techniky založené na strojovom učení
 - ▣ Informačné značky sú vytvárané deterministickými algoritmami
 - ▣ Možnosť naučiť sa závislosti medzi zmenami označovaného obsahu a potrebnej údržby informačných značiek
- Prístupy založené na využití „davu“
 - ▣ Úspešne použité pri iných zložito riešiteľných problémoch
 - ▣ Napr. využitie implicitnej spätnej väzby na určenie správnosti ukotvenia značky a platnosti obsahu značky
- ICWE 2012 DC, LNCS 7703

Model informačných značiek

- Založený na Open Annotation Collaboration
 - ▣ Prepracovaný do objektovej štruktúry
 - ▣ Použité len potrebné časti
 - ▣ Spätne kompatibilný
- Ukladaný v dokumentovej databáze MongoDB
 - ▣ Možnosť analýzy veľkého množstva dát
 - ▣ Navrhnutý algoritmus vyhodnocovania SPARQL
- WISM 2012, LNCS 7518

Zhrnutie

- Problém: Ako reprezentovať a udržiavať metadáta dynamických informačných systémov?
- Navrhnutá 1. verzia modelu a úložiska informačných značiek
- Plán
 - ▣ Zber informačných značiek a štúdium ich vlastností
 - ▣ Overovanie rôznych prístupov údržby pri rôznych typoch informačných značiek
 - Porovnávaním nanovo vygenerovaných informačných značiek a automaticky opravených informačných značiek

Model informačných značiek

