

ROSES Really Open Simple and Efficient Syndication

RSS

Ressources web

- Contenu riche
- Utilité à long-terme
- Mode « pull »

Flux RSS

- Résumés simples
- Utilité à court-terme
- Mode « push »

Début – fin : 1 janvier 2008 – 31 décembre 2010

Site Web : <http://www-bd.lip6.fr/roses>

Gforge : <http://gforge.cnam.fr/gf/project/roses/>

Contact Bernd.Amann@lip6.fr

Montant alloué : 294 k€

Partenaires :

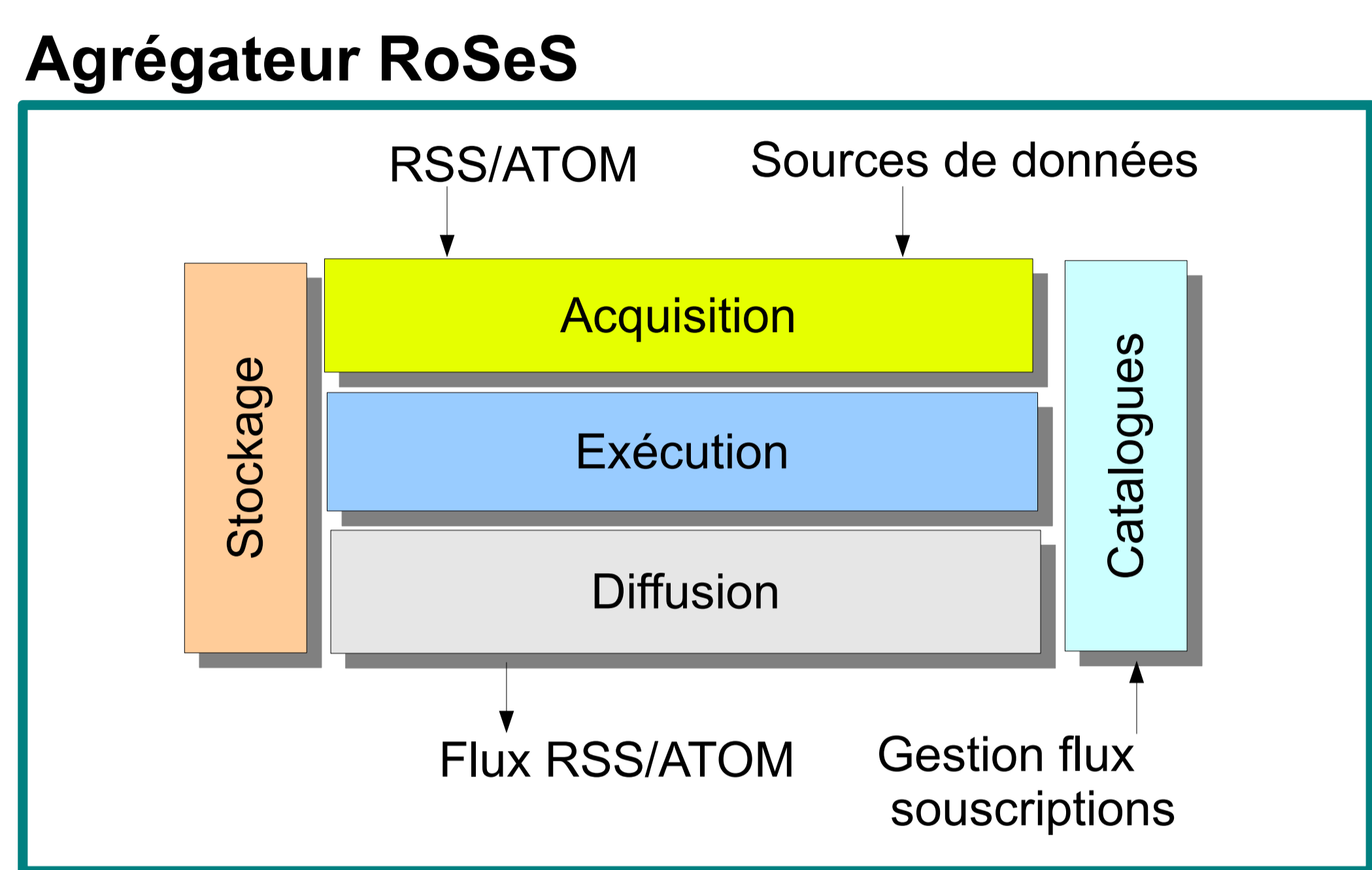
- LIP6-UPMC (Wisdom)
- Cedric-CNAM (Wisdom)
- PRISM – UVSQ
- LSIS – Univ. de Toulon
- 2or3things

Cas d'applications

- Actualités, news
- Médias sociaux (blogs, Twitter)
- Surveillance de sites web
- Objets mobiles
- Analyse de série temporelles (bourse)

Passage à l'échelle

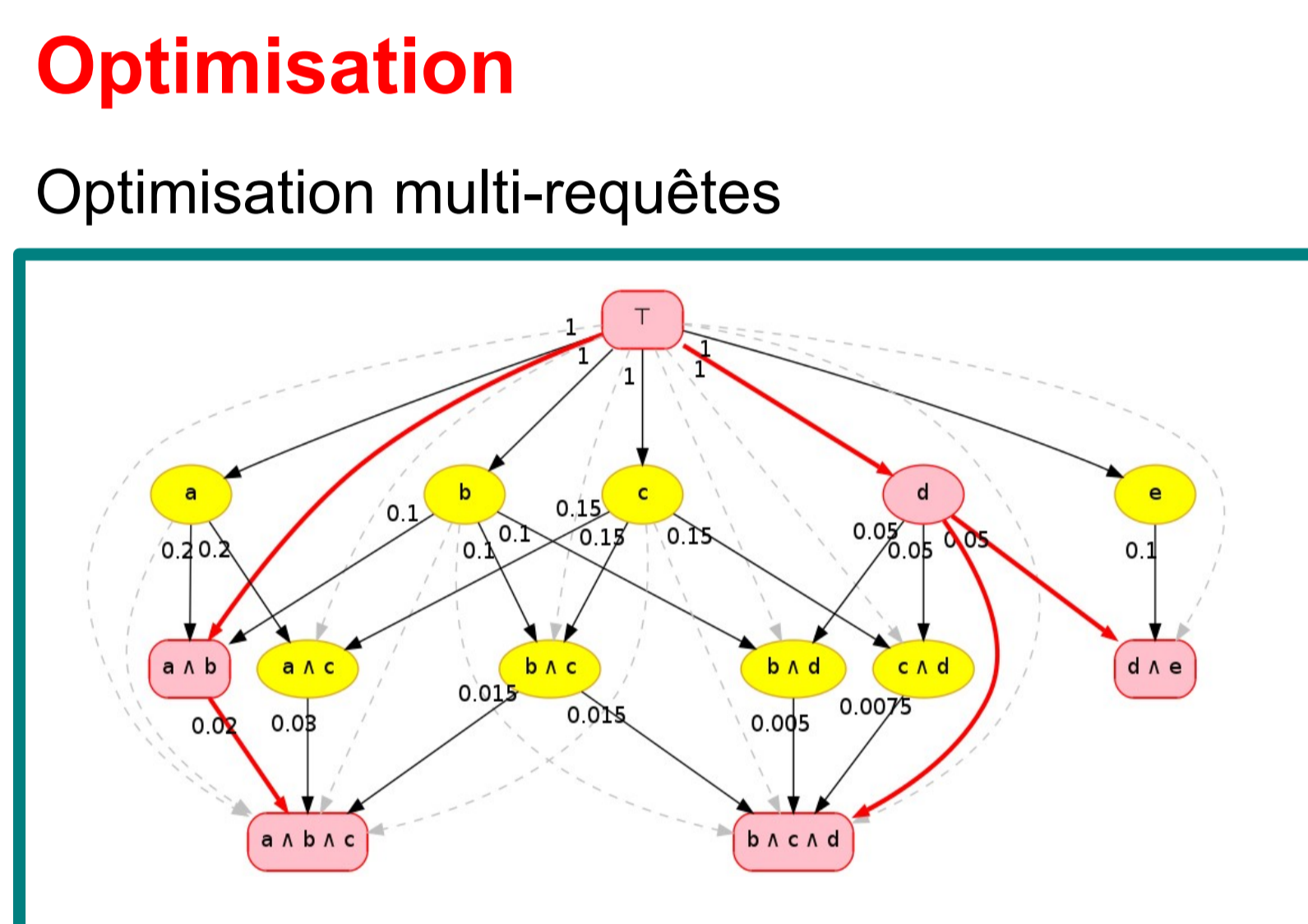
- Optimisation multi-requêtes
- Indexation de souscriptions
- Stratégies de rafraîchissement RSS
- Caches distribués P2P
- Ranking et requêtes top-k continues
- Bancs d'essais



Langages et algèbres

XQuery en « continu »

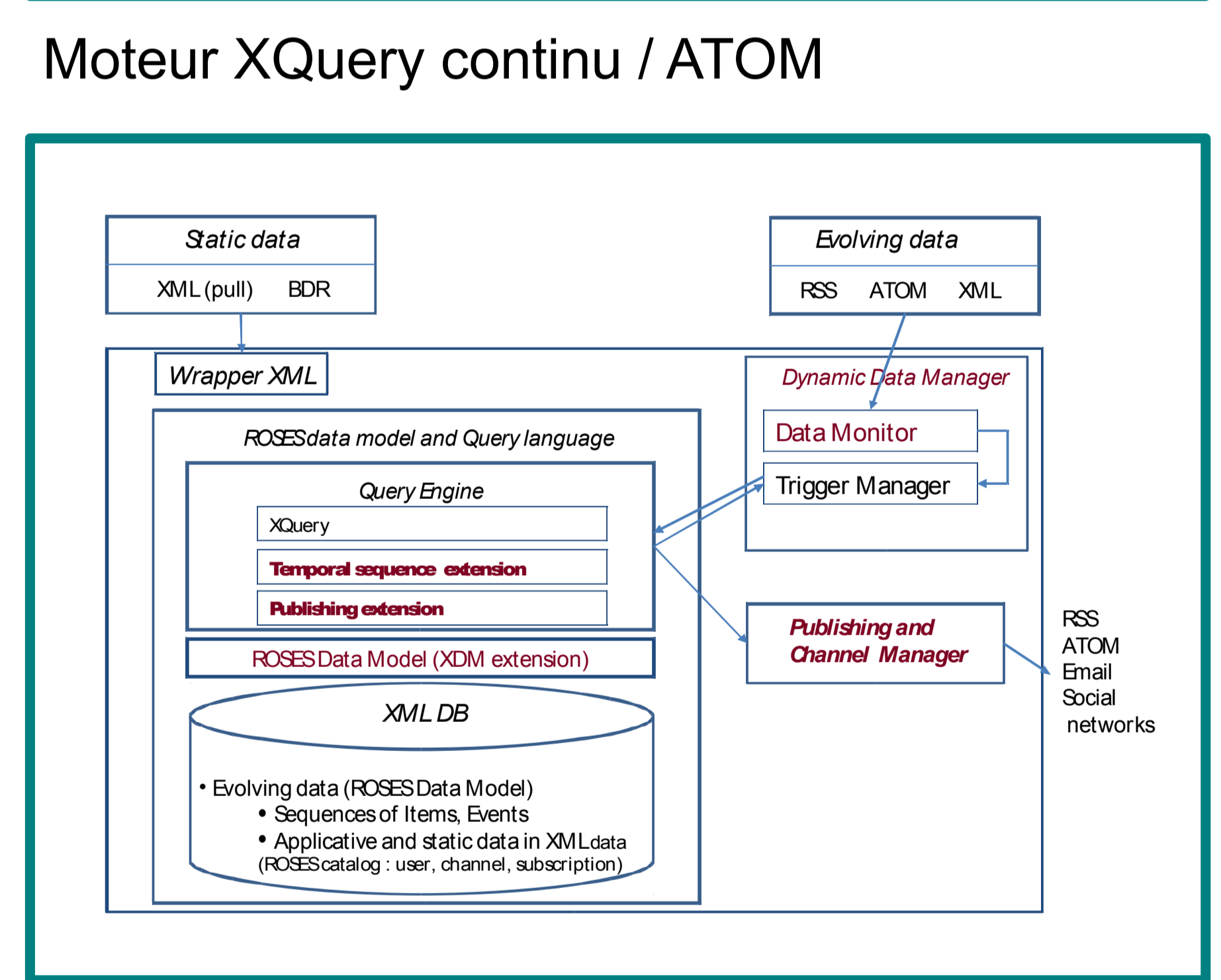
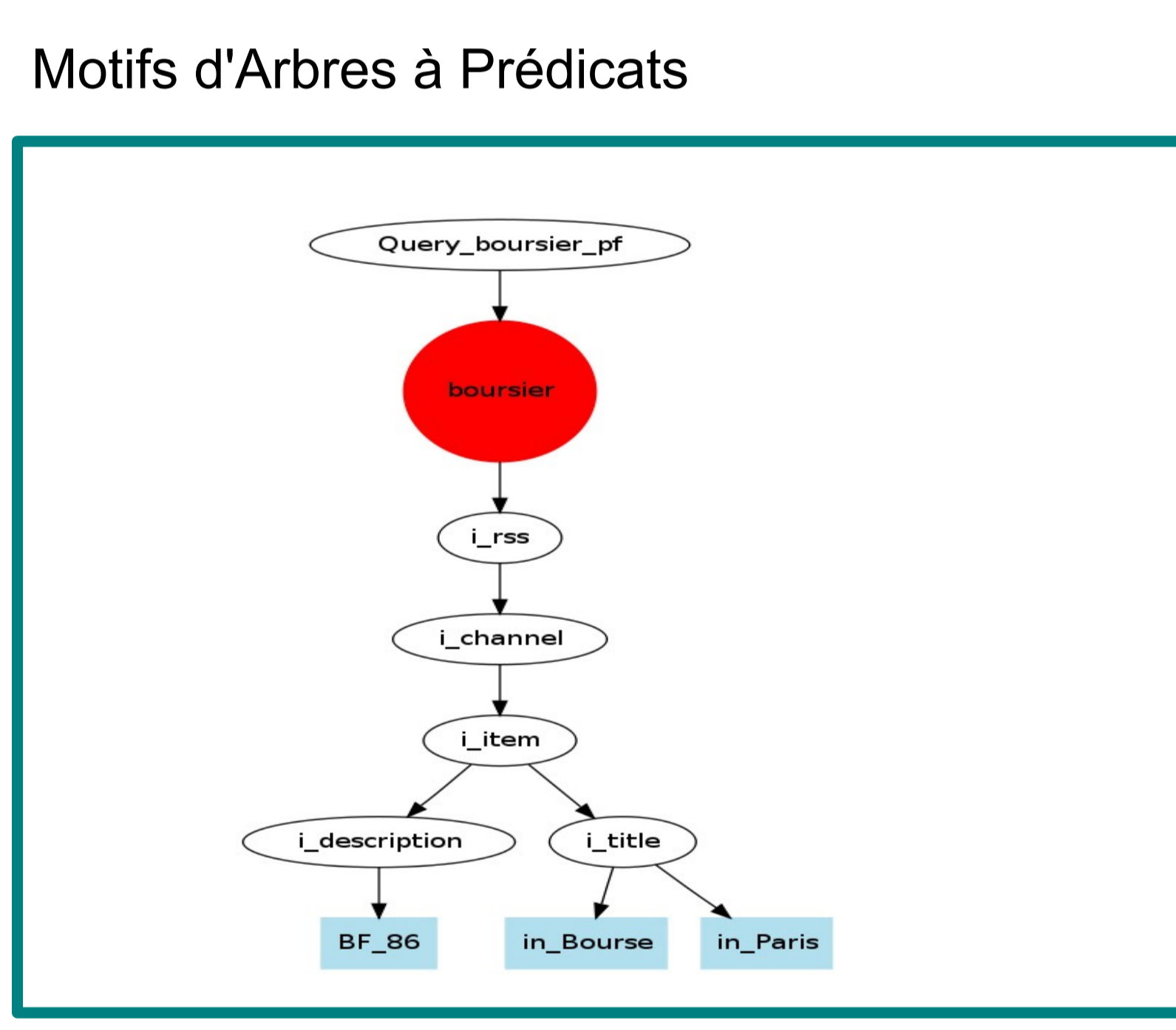
```
for $i in feed(tumbling, 1, new, 'boursier')/rss/channel/item,
    $pf in /data/portfeuille
where contains($i/title, 'Bourse') or contains($i/title, 'Paris')
and contains($i/description, $pf/action/nom)
return <a_lire>{$i}</a_lire>
```



Prototypes

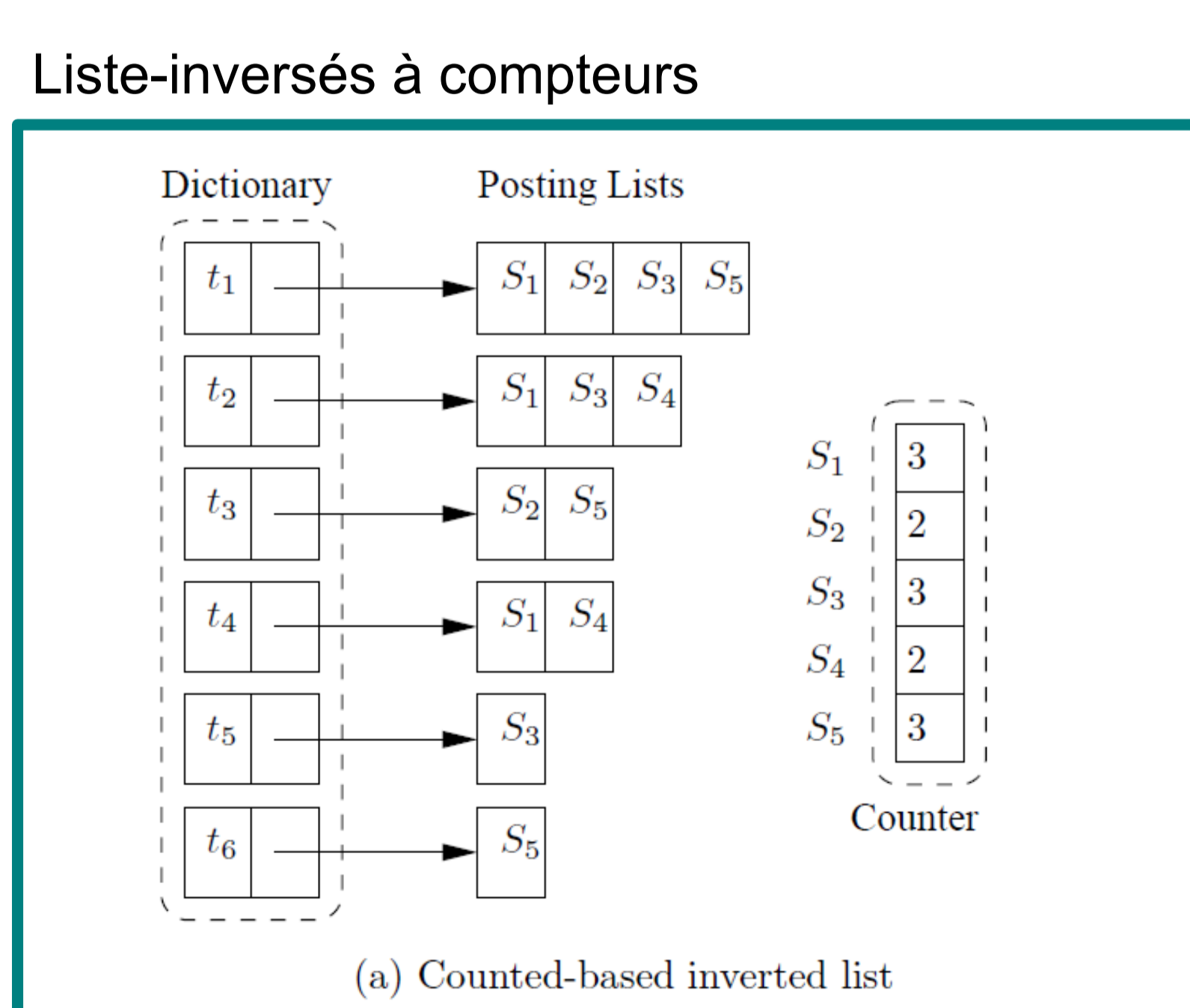
- Aggrégateur ROSES (CNAM, LIP6)
- Extension XQuery : Triggers et Fenêtres
- Crawler RSS
- Générateur de requêtes
- Emulateur de flux RSS

Algèbre de flux RSS



Analyse et Banc d'Essai

Type	% feeds	% items	ratio
Social Media	1.77%	9.45%	7085
Press	9.99%	38.82%	5141
Forum	1.51%	3.62%	3178
Sales	11.32%	15.49%	1811
Misc.	41.47%	25.47%	812
Blog	33.93%	7.14%	278



Interfaces Utilisateurs

Publications (sélection)

- RoSeS: A Continuous Query Processor for Large-scale RSS Filtering and Aggregation. J. Creus, B. Amann, N. Travers, D. Vodislav, démonstration, CIKM 2011.
- RoSeS: A continuous content-based query engine for RSS feeds, J. Creus, B. Amann, N. Travers, D. Vodislav, DEXA 2011.
- Characterizing Web Syndication Behavior and Content. Z. Hmedeh, N. Travers, N. Vouzoukidou, V. Christophides, C. du Mouza, M. Scholl, WISE 2011.
- Online Refresh Strategies for RSS Feed Crawlers, R. Horincar, B. Amann, T. Artières, BDA 2011.
- Best-effort refresh strategies for content-based RSS feed aggregation, R. Horincar, B. Amann and T. Artières, WISE 2010.
- A Semantic Map of RSS Feeds to support Discovery, G. Hochard, Z. Lacroix, J. Creus et B. Amann, RED/iiWAS 2010.
- Efficient P2P Processing of Times Series: Application to Stock Investment and Mobile Objects Analysis. G. Gardarin, B. Nguyen, L. Yeh., et.al., BDA 2009.