

Les entités nommées

Alexandre Gefen
(CNRS-Université Paris 4-Sorbonne)

Qu'est qu'une entité nommée ?

— [Les unités nommées sont des unités lexicales sélectionnées pour leur intérêt sémantique : dates, unités, noms de lieux, etc. Elles sont à la fois difficilement compréhensibles en contexte et en même temps comparables à d'autres éléments similaires.

Définition linguistique

— [Définition (via N. Dufournaud) :

— [– « Tous les éléments du langage définis par référence : les noms propres au sens classique, les noms propres dans un sens élargi mais aussi les expressions de temps et de quantité » (Friburger, 2006).

NB vient de Kripke (désignateurs rigides)

Les noms

— [Quatre types :

- Personnes
- Lieux
- Les fonctions
- Les organismes

Les valeurs



Dates, heures



Montants



Numérotations

Pourquoi s'intéresser aux EN?

- **Utiliser les textes comme une base données :**

- [Mise en relation des textes

- [Traitement type TAL, cartographie, index (facile avec XSLT)

- [Suivi de tendance, veille

- [Optimisation de recherche et moteur de recherche sémantique

- [Production de savoir par inférence (classification de documents par exemple)

Utile avec par exemple :

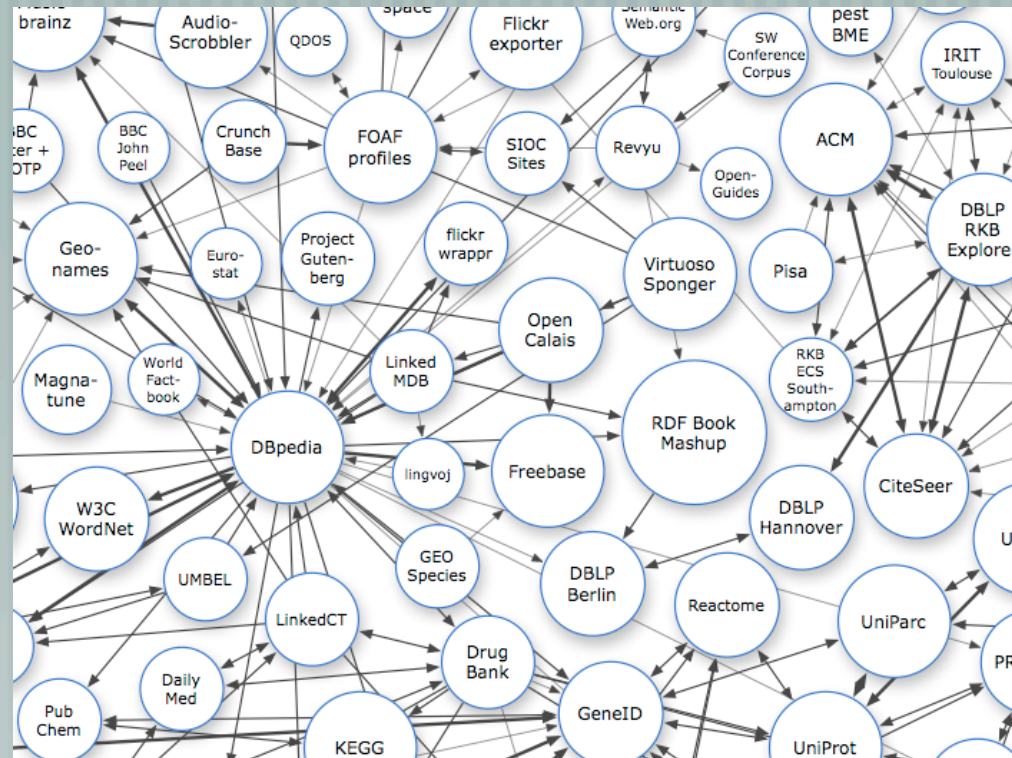
- [des dictionnaires

- [des textes de références

- [des textes très longs et complexes : ex.

- [Le Grand Cyrus partie 1, livre 1](#)

En perspective : le web sémantique



Le balisage des textes vise à les faire communiquer entre eux mais aussi à créer un web de données destiné à s'intégrer au web sémantique

Exemples

— [Navigateur ontologique basée sur SKOS : [YAGO 2](#)

— [Fiches automatisée de Freebase : [Henry Ford facts - Freebase](#)

— [DBpedia : extraction de Wikipedia :
[Exemples SPARQL - Dbpedia.fr](#)

Présupposé de la production d'EN

- un travail d'identification (de balisage) manuel ou dans certains cas automatique
- un travail d'indexation et de documentation
- une catégorisation (arbre de connaissance ou même une ontologie)

Les problèmes de sémantisaion

— [On cherche à définir la référence unique de l'entité que l'on va nommer : différence avec le niveau sémantique : Alexandre Gefen=deux mots mais une seule entité

— [Problèmes : orthographe et variantes, identité sociale, pseudonyme, limite imprécise entre les titres et les fonctions, changement de noms de lieux (Lutèce/Paris) ou de système de datation

Exemple classique

- L'ancien premier ministre socialiste Lionel Jospin a confirmé, jeudi 28 septembre, sur RTL, qu'il ne sera pas candidat à l'investiture socialiste pour la présidentielle de 2007.



- L'ancien premier ministre socialiste <pers>Lionel Jospin</pers> a confirmé, <date>jeudi 28 septembre</date>, sur <org>RTL</org>, qu'il ne sera pas candidat à l'investiture socialiste pour la présidentielle de <date> 2007</date>.

Par delà les noms propres, gestion de grands corpus d'information en intégrant les fonctions, les organisations

- L'ancien premier ministre socialiste



- L'ancien `<role>premier ministre</role>`
`<org>socialiste</org>`

Dans la TEI

Dans la TEI : tout encoder avec `<rs>` balise générique

- `<p>`La famille s'était alors retirée en banlieue, à `<rs type="place">`Villemomble`</rs>`, mais `<rs type="person">`Alfred `</rs>`aimait se rendre à `<rs type="place">`Paris`</rs>` et `<rs type="date">`un jour de 1917`</rs>` alors qu'il sortait de `<rs type="place">`la maison des Arts et Métiers`</rs>` il était tombé inanimé dans la rue.`</p>`

Des balises spécialisées

— [<name> pour les noms propres et <persName> pour les personnes (regroupés en <listPerson> mais non imbriqués)

— [<orgName> pour les institutions (regroupés en <listOrg> et éventuellement imbriqués)

— [<place> pour les lieux (regroupés en <listPlace> et éventuellement imbriqués)

— [<date> pour les dates (avec attribut par exemple : @when pour la précision, @from @to pour un intervalle de temps certain, @notBefore @notAfter pour un intervalle de temps incertain, @calendar pour les calendriers)

Détails de <persName>

- <surname> nom de famille ou patronyme

- <forename> prénom

- <genName> generational name

- <roleName> titre et avant-nom

- <addName> épithète

- <nameLink> particules nobiliaires

-

Le cas de PersName

PersName

— [Très utile par exemple dans le TEI Header

— [Peut être utilisé aussi dans le corps du texte

— [Peut être limité à une référence externe sans sémantisation :

- `<persName ref="http://fr.wikipedia.org/wiki/Alexandre_Gefen">Alexandre Gefen</persName>`

Exemple avec <persName>

[<persName>

[<forename>Edward</forename>

[<forename>George</forename>

[<surname>Bulwer-Lytton</surname>, <roleName>Baron Lytton of

[<placeName>Knebworth</placeName>

[</roleName>

[</persName>

Exemple 2 avec des type (via N. Dufournaud)

- ```
<persName>
<addName type="epithete-honneur">haute dame</addName>
<roleName type="avant-nom">dame</roleName>
<forename>Suzanne</forename>
<nameLink>de</nameLink>
<surname type="maison">Bourbon</surname>
<roleName type="titre">dame</roleName>
<nameLink>de</nameLink>
<placeName type="terre">Rieux</placeName>

</persName>
```
-

# De la balise à la référence, deux outils de pointage par attribut

Par un code avec l'attribut @key= "moncode" permet de regrouper des entités nommées :

```
<persName key= "11" >Lou Burnard</persName> ...
```

```
<persName key= "11" >L. Burnard</persName>
```

# Exemple de définition externe

- Pour pointer vers une définition : @ref= "maref " :

- `<persName ref="#AG">Alex</persName>....`

- `<persName ref="#AG">Alexandre Gefen</persName> ....`

- `<persName ref="#AG">M. Gefen</persName>`

- Le détails se met dans `<person>` :

- `<person xml:id="AG">`

- `<persName xml:lang="fr">`

- `<persName>Alexandre Gefen</persName>`

- `<persName> <forename>Alexandre</forename> <surname>Gefen</surname> </persName>`

- `<birth when="1970"/>`

- `<note>Alexandre est chercheur ... </note>`

- `</person>`

# Exemple de liste

- `<listPerson>`  
`<person xml.id="AG">`
- `<persName>Alexandre Gefen</persName>`  
`</person>`  
`<person xml.id="LB">`
- `<persName>Lou Burnard</persName>`  
`</person>`  
`</listPerson>`



# Exemple de relation

- `<listPerson>`  
`<person xml.id="AG">`
- `<persName>Alexandre Gefen</persName>`  
`</person>`  
`<person xml.id="LB">`
- `<persName>Lou Burnard</persName>`  
`</person>`
- `<relation name="travail" mutual="#AG #LB"/>`  
`</listPerson>`

# D'innombrables possibilités...

— [ Exemples exhaustifs (du dictionnaire à l'index en passant par la bibliographie) :

Example: <persName> (personal name)

— [ D'autres présentations meilleures que la mienne :

Data vs Text: encodage des entités non-textuelles ou

[www.bvh.univ-tours.fr/actualites/TEI\\_index\\_11-02-01.pdf](http://www.bvh.univ-tours.fr/actualites/TEI_index_11-02-01.pdf)

# Exercice

- [ Faire votre fiche à l'intérieur d'un texte bref inventé en utilisant `<persName>`
- [ Extraire votre définition avec `@ref` et `<person>`
- [ Ajouter des relations familiales en utilisant `<relation>` et les attributs `@name`, `@type`, `@active`, `@mutual`, `@passive`
- [ Ajouter des événements biographiques en utilisant `<event>`

# Merci !

[alexandre.gefen@paris-sorbonne.fr](mailto:alexandre.gefen@paris-sorbonne.fr)