

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

# Données et défis

## Exposé niveau lycée

Matthieu Objois   Pierre Senellart



8 février 2005

# Beaucoup de données

## Données et défis

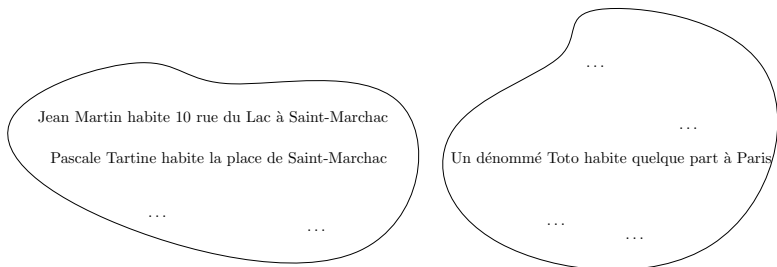
Matthieu  
Objois,  
Pierre  
Senellart

## Données

## Temporalité

## Web caché

## Conclusion



- **Beaucoup** de données (plusieurs millions d'adresses!).
- Mais jamais besoin de tout en même temps, on veut juste **poser des questions**.

# Beaucoup de données

## Données et défis

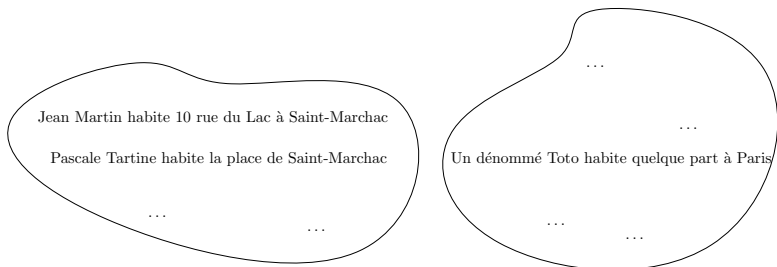
Matthieu  
Objois,  
Pierre  
Senellart

## Données

## Temporalité

## Web caché

## Conclusion



- **Beaucoup** de données (plusieurs millions d'adresses!).
- Mais jamais besoin de tout en même temps, on veut juste **poser des questions**.

# Où habite Jean Martin ?

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

Nom

Prénom

Adresse

ex : 7 av de la cristallerie

Localité

Guide

ex : rennes, paris 9, 37100

Département  
ou région

Guide

ex : 92, finistère, aquitaine, idf

Rechercher

# Où habite Jean Martin ?

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

**Nom**

**Prénom**

**Adresse**

ex : 7 av de la cristallerie

**Localité**

ex : rennes, paris 9, 37100

Guide

**Département  
ou région**

ex : 92, finistère, aquitaine, idf

Guide

Rechercher

# Où habite Jean Martin ?

## Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

## Données

## Temporalité

## Web caché

## Conclusion

Nom martin

Prénom jean

Adresse

ex : 7 av de la cristallerie

Localité saint-marchac

ex : rennes, paris 9, 37100

Guide

Département  
ou région

ex : 92, finistère, aquitaine, idf

Guide

Rechercher

« ET »

# Question plus compliquée

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

## Question

*Je connais quelqu'un qui s'appelle Martin à Saint-Marchac, mais je ne sais plus si c'est son nom ou son prénom. Comment le retrouver ?*

Impossible en une seule fois avec les Pages Blanches !

Il faudrait poser une question dans un nouveau langage :

(Prénom = Martin OU Nom = Martin) ET Ville = Saint-Marchac

# Question plus compliquée

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

## Question

*Je connais quelqu'un qui s'appelle Martin à Saint-Marchac, mais je ne sais plus si c'est son nom ou son prénom. Comment le retrouver ?*

**Impossible** en une seule fois avec les Pages Blanches !

Il faudrait poser une question dans un nouveau langage :

(Prénom = Martin **OU** Nom = Martin) **ET** Ville = Saint-Marchac

# Question plus compliquée

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

## Question

*Je connais quelqu'un qui s'appelle Martin à Saint-Marchac, mais je ne sais plus si c'est son nom ou son prénom. Comment le retrouver ?*

**Impossible** en une seule fois avec les Pages Blanches !

Il faudrait poser une question dans un nouveau **langage** :

(Prénom = Martin **OU** Nom = Martin) **ET** Ville = Saint-Marchac

## Question encore plus compliquée

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

- On peut encore imaginer des questions plus compliquées. . .  
*Quelles sont les villes dans lesquelles il n'y a pas deux personnes avec le même prénom ?*

**Impossible** dans le langage précédent ! Il faudrait encore inventer un nouveau langage.

- L'ordinateur va mettre plus de temps pour répondre à la question : **pour chaque** ville, il va falloir comparer 2 à 2 les prénoms de ses habitants. . .

### Exemple

Si 100 villes ont chacune 1000 habitants, il faut (naïvement) :

$$100 \times \frac{1000 \times (1000 - 1)}{2} \approx 50\,000\,000$$

comparaisons.

## Question encore plus compliquée

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

- On peut encore imaginer des questions plus compliquées. . .

*Quelles sont les villes dans lesquelles il n'y a pas deux personnes avec le même prénom ?*

**Impossible** dans le langage précédent ! Il faudrait encore inventer un nouveau langage.

- L'ordinateur va mettre plus de temps pour répondre à la question : **pour chaque** ville, il va falloir comparer 2 à 2 les prénoms de ses habitants. . .

### Exemple

Si 100 villes ont chacune 1000 habitants, il faut (naïvement) :

$$100 \times \frac{1000 \times (1000 - 1)}{2} \approx 50\,000\,000$$

comparaisons.

## Question encore plus compliquée

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

- On peut encore imaginer des questions plus compliquées. . .  
*Quelles sont les villes dans lesquelles il n'y a pas deux personnes avec le même prénom ?*

**Impossible** dans le langage précédent ! Il faudrait encore inventer un nouveau langage.

- L'ordinateur va mettre plus de temps pour répondre à la question : **pour chaque** ville, il va falloir comparer 2 à 2 les prénoms de ses habitants. . .

### Exemple

Si 100 villes ont chacune 1000 habitants, il faut (naïvement) :

$$100 \times \frac{1000 \times (1000 - 1)}{2} \approx 50\,000\,000$$

comparaisons.

# Compromis expressivité — efficacité

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

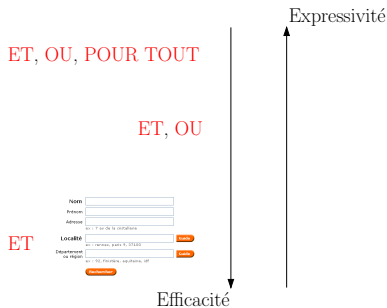
Données

Temporalité

Web caché

Conclusion

- Langage  $\rightarrow$  ensemble de questions qu'on peut poser (**expressivité** du langage)
- Plus un langage est **expressif** (plus on peut poser de questions avec), plus **les temps de réponse** sont importants...



# Ajouter du temps aux données

Données et  
défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

Sur un annuaire, on ajoute des informations de temps :

- Jean Martin habitait Saint-Marchac en 1935.
- Toto habite Paris depuis le 3 janvier 1982.
- ...

# Ce que je fais (Matthieu)

Données et  
défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

- Il y a des **langages** pour poser des questions sur les données avec du temps.
- **Comparer** leur expressivité.
- **Déterminer** leur efficacité.
- Et en inventer de nouveaux !

# Le Web (caché), une source de données

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

- **Beaucoup** de données sur le World Wide Web (plusieurs milliards de pages!)
- Données localisées à différents endroits.
- Web caché :

## Definition (Web caché)

Ensemble des pages Web qui ne sont pas accessibles par des liens.

- Exemple : Pages Blanches, horaires de train...
- On estime qu'il y a **500 fois plus** de données sur le Web caché que sur le Web « visible ».

# Ce que je fais (Pierre)

Données et défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

- **Trouver** et **décrire (automatiquement)** les différents services du Web caché (Pages Blanches, calculs d'itinéraires...)
- **Répondre à des questions**, en demandant aux différents services nécessaires et en **combinant automatiquement** leurs réponses :

*Combien me faut-il de temps pour aller chez Jean Martin ?*

- *Où habite Jean Martin ?*

→ Pages Blanches



- *Combien de temps pour aller de Paris à Saint-Marchac ?*

→ Mappy

# Conclusion

Données et  
défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

- Un aspect aperçu des bases de données : **questions**, **expressivité**, **efficacité**
- D'autres aspects intéressants :
  - **Organiser** de manière appropriée les données.
  - Ajouter, Modifier, Supprimer **efficacement** des données.
  - ...

# Conclusion

Données et  
défis

Matthieu  
Objois,  
Pierre  
Senellart

Données

Temporalité

Web caché

Conclusion

- Un aspect aperçu des bases de données : **questions**, **expressivité**, **efficacité**
- D'autres aspects intéressants :
  - **Organiser** de manière appropriée les données.
  - Ajouter, Modifier, Supprimer **efficacement** des données.
  - ...