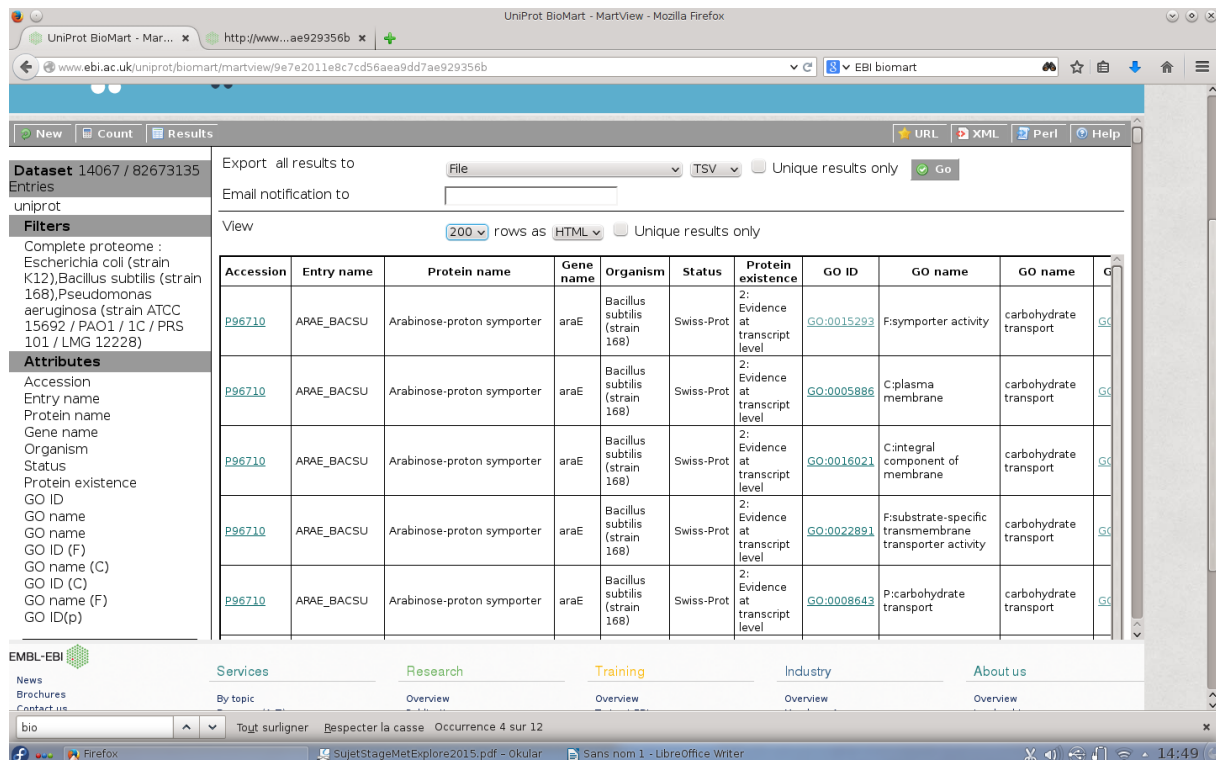


Compte rendu du TP Biomart

Maxime CHAZALVIEL



UniProt BioMart - MartView - Mozilla Firefox

http://www.ebi.ac.uk/uniport/biomart/martview/9e7e2011e8c7cd56aea9dd7ae929356b

Dataset 14067 / 82673135
Entries uniprot

Filters
Complete proteome : Escherichia coli (strain K12), Bacillus subtilis (strain 168), Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228)

Attributes
Accession
Entry name
Protein name
Gene name
Organism
Status
Protein existence
GO ID
GO name
GO ID (F)
GO name (C)
GO ID (C)
GO name (F)
GO ID(p)

Export all results to File TSV Unique results only Go

Email notification to

View 200 rows as HTML Unique results only

Accession	Entry name	Protein name	Gene name	Organism	Status	Protein existence	GO ID	GO name	GO name
P96710	ARAE_BACSU	Arabinose-proton symporter	araE	Bacillus subtilis (strain 168)	Swiss-Prot	2: Evidence at transcript level	GO:0015293	F:symporter activity	carbohydrate transport
P96710	ARAE_BACSU	Arabinose-proton symporter	araE	Bacillus subtilis (strain 168)	Swiss-Prot	2: Evidence at transcript level	GO:0005886	C:plasma membrane	carbohydrate transport
P96710	ARAE_BACSU	Arabinose-proton symporter	araE	Bacillus subtilis (strain 168)	Swiss-Prot	2: Evidence at transcript level	GO:0016021	C:integral component of membrane	carbohydrate transport
P96710	ARAE_BACSU	Arabinose-proton symporter	araE	Bacillus subtilis (strain 168)	Swiss-Prot	2: Evidence at transcript level	GO:0022891	F:substrate-specific transmembrane transporter activity	carbohydrate transport
P96710	ARAE_BACSU	Arabinose-proton symporter	araE	Bacillus subtilis (strain 168)	Swiss-Prot	2: Evidence at transcript level	GO:0008643	P:carbohydrate transport	carbohydrate transport

EMBL-EBI

News
Brochures
Contact us

Services
By topic
Overview

Research
Overview

Training
Overview

Industry
Overview

About us
Overview

bio

Tout surligner
Respecter la casse
Occurrence 4 sur 12

SujetStageMetExplore2015.pdf - Okular

Sans nom 1 - LibreOffice Writer

14:49

Introduction

Biomart est un outil simple pour la recherche et le traitement de données biologiques. Cet outil a pour but d'encourager la collaboration scientifique en proposant le format d'échange XML utilisable par tous. Ce format permet de faire face aux différences entre les machines utilisées par les scientifiques. Biomart propose également un large choix de banques de données (Ensembl, Interpro, Uniprot) .Et lors du TP nous avons utilisé Uniprot.

Mode d'emplois

Les différents filtres permettent de trier l'information plus précisément, et de façon plus pertinente pour obtenir seulement les données nécessaire. Il est donc possible de récupérer le nom des protéines, des gènes, leurs identifiants de Gene Ontology sur la fonction de la protéine ou ceux du processus biologique dans lequel elle pourrait être impliquée.

Il est aussi possible de récupérer les identifiants appartenant à d'autres bases de données référencées.

Intégration

Sources de données distantes

Pour mettre à disposition des données distantes depuis notre serveur, nous devons installer Biomart. Il faut de compiler le fichier java présent dans le dossier téléchargé avec le logiciel Ant.

Une fois sur l'interface de Biomart il est nécessaire d'ajouter les sources sur lesquels l'utilisateur veut travailler. Lors du TP nous avons encore utilisé Uniprot.

Il faut ensuite démarrer le serveur par l'interface graphique ou dans un terminal avec la commande `./dists/scripts/biomart-server.sh start`.

Biomart démarre sur le navigateur Web de l'ordinateur en utilisant le port local 9000 et affiche une interface identique au site web.

Source de données locale

configuration

Martconfigurator donne le choix entre deux options lors de l'ajout de sources. La première est celle que nous avons utilisé dans le TP pour le serveur distant est URL et permet l'utilisation de banque telle qu'Uniprot. La seconde, est Relationnal Database Management System RDBMS elle permet d'utiliser des bases de données, chargée en amont avec mysql par exemple, sur un serveur local.

Ensuite il faut saisir les paramètres de connexion au serveur de base de données (port 3306) et sélectionner la base restaurée auparavant.

Pour finir il faut faire de même que pour les serveurs distants.

Edition de lien entre source et local

L'édition de lien entre source et local permet de réaliser des jointures entre les tables de chacune de ces bases de données. Nous pouvons donc faire correspondre des attributs de la base Uniprot à la base de données chargée en local. Il est ainsi possible d'obtenir les numéros d'accension des gènes, les termes de la Gene Ontologie.

Grace à la page suivante nous pouvons obtenir les commandes d'accès au serveur au format permettant de gérer les attributs des bases de données :
<http://localhost:9000/martsoap?wsdl>

La liste Marts au format XML est également disponible ainsi qu'une documentation sur les filtres.

La création d'un script python permettant de récupérer et de lister les fonctionnalités au format XML permet aux bioinformaticien d'automatiser les requêtes sur les banques de données pour faciliter l'accès contenu.

Conclusion

Je pense que le principal intérêt de Biomart est le fait que l'on puisse l'utiliser sur sa machine personnelle. Le format XML permet son utilisation dans n'importe quels logiciels utilisé par les chercheurs