REMORA: un poisson pilote dans l'océan des web-services BioMOBY

Sébastien Carrere [1], Jérôme Gouzy[1]
[1] Laboratoire des Interactions Plantes-Microorganismes INRA/CNRS
Sebastien.Carrere@toulouse.inra.fr, Jerome.Gouzy@toulouse.inra.fr
http://bioinfo.genopole-toulouse.prd.fr/remora

Emerging web-services technology allows interoperability between multiple distributed architectures. Here, we present REMORA, a web server implemented according to the BioMoby web-service specifications, providing life science researchers with an easy-to-use workflow generator and launcher, a repository of predefined workflows and a survey system.

Résumé

En bioinformatique comme dans d'autres domaines scientifiques, l'intéropérabilité des systèmes informatiques est devenu un élément clé pour accéder non seulement à la masse d'informations mais également aux outils qui permettront de la traiter et de l'intégrer. Durant ces dernières années, deux technologies ont émergé parallèlement pour répondre à ces questions: les grilles de calcul ou de données d'une part, les web-services d'autres part. Cependant, on assiste actuellement à une convergence de ces deux technologies mais que ce soit du coté des grilles de calcul que des web services, de nombreux problèmes techniques demeurent (exécution asynchrone, robustesse, fiabilité, maintenabilité des données, authentification, etc..) et le choix des standards, au-delà de SOAP, n'est pas encore figé. Mais même si l'on peut espérer que des solutions normalisées émergent rapidement, l'utilisation de ces technologies ne se justifiera vraiment en bioinformatique que lorsque les biologistes pourront accéder de façon intuitive à ces ressources informatiques pour récupérer, analyser et intégrer les données nécessaires à leurs recherches et ce avec le maximum de transparence et de fiabilité. Or, répondre à ce besoin nécessite le développement d'interfaces utilisateurs adaptées car malheureusement, les outils d'ores et déjà disponibles comme Taverna [1] sont destinés principalement à des utilisateurs informaticiens et, pour rester génériques, sont relativement complexes à mettre en œuvre pour un utilisateur non programmeur. C'est à la suite de ce constat que nous avons récemment développé le serveur web REMORA [2]. Dans un premier temps, REMORA tire parti du typage, avec sémantique bioinformatique, des entrées sorties des web-services BioMOBY [3] pour permettre la découverte des ressources, puis via une interface la plus simplifiée possible, rend possible la génération étape par étape et l'exécution de chaînes de traitement complexes. De plus, afin de partager l'effort de développement, REMORA autorise aussi bien l'enregistrement sur le site des workflows les plus fréquemment utilisés mais également, pour les projets pour lesquels la veille est critique, la ré-exécution périodique des chaînes de traitement.

Web-services

Technologies

myGrid
programme *e-Science* (UK)
messages WSDL

BioMOBY
approche collaborative
XML avec typage bioinformatique encapsulé dans du WSDL
utilisé par une partie de la communauté bioinformatique “Plante” (Prg EU: Planet, GLIP)

Développements

Code
Bibliothèque utilisant **XSL** pour la manipulation de gros messages XML
Procédure de développement de web-services BioMOBY::Perl

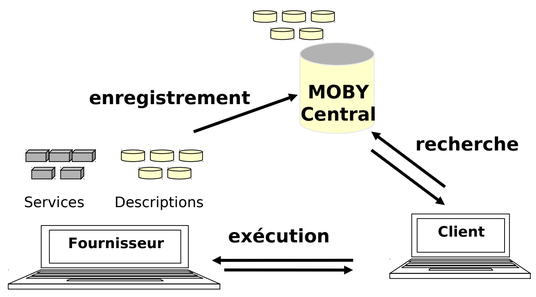
Services
68 services à disposition de la communauté dont 33 services **EMBOSS**

BioMOBY

La mise à disposition d'un web-service BioMOBY se déroule en deux étapes.

1) développement du service sur un serveur accessible via le web (**SOAP**). Il existe trois API: **JAVA**, **Perl** et **Python**.

2) enregistrement du service dans l'annuaire (ou central). La validation de l'enregistrement se fait via la restitution d'un fichier **RDF**.



Grâce à **l'ontologie** des services et de leurs interfaces, BioMOBY permet la découverte de nouvelles ressources. **L'annuaire** retourne au client la description des services correspondants (fournisseur, paramètres d'appel, entrées, sorties).

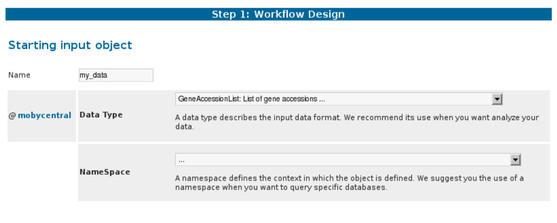
L'exécution du web-service se fait via le protocole SOAP, le client se connectant au ***dispatcher*** mis à disposition par le fournisseur. Le message transmis contient les données d'entrées et les paramètres du web-service.

Le service exécuté, un message SOAP est retourné au client. Ce message contient un document **XML BioMOBY** incluant le résultat produit par le service.

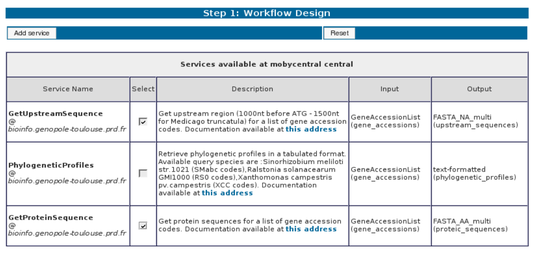
Depuis Septembre 2006, l'API BioMOBY décrit l'exécution de services **asynchrones** via le protocole **WSRF**.

Workflows

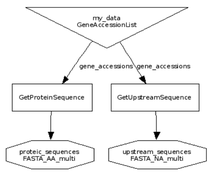
Démarrage



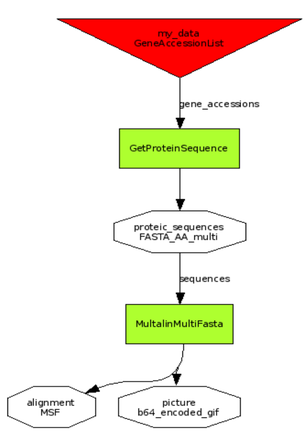
Le point de départ de la construction s'appuie sur le typage bio-informatique des interfaces. Ainsi l'utilisateur choisit dans une liste le type de données qu'il souhaite analyser.
Le système se charge alors de proposer une liste de services applicables. Ces services sont ordonnés par pertinence décroissante, du service prenant en entrée le type de données sélectionné au service le plus généraliste, tout en respectant l'ontologie des interfaces.



L'utilisateur sélectionne les services qu'il souhaite exécuter et une représentation graphique du workflow est générée.

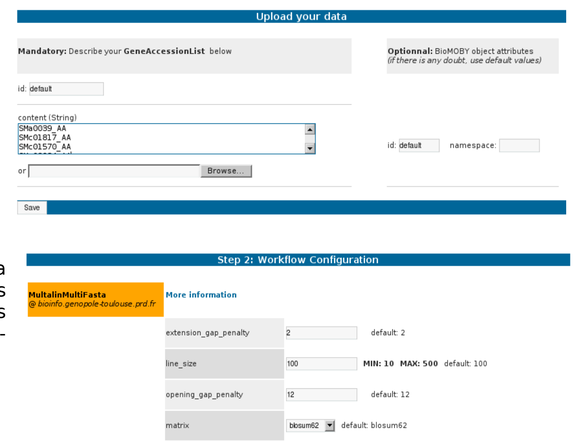


Configuration

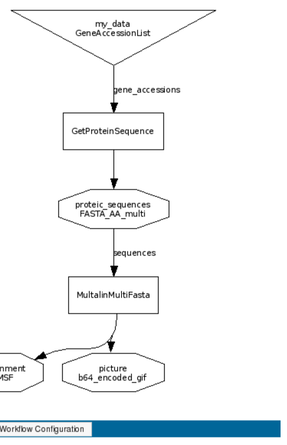


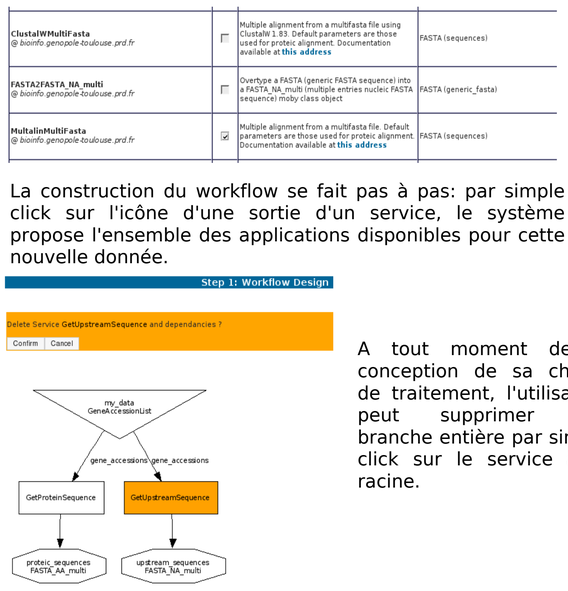
Un code couleur simple permet de signaler à l'utilisateur les étapes bloquantes, notamment les données à fournir pour initier le workflow. Ceci peut être fait via un formulaire par simple copier/coller de données ou chargement de fichier.

REMORA exploite également la notion d'*articles secondaires* des services, qui sont les paramètres facultatifs d'exécution du web-service.



Construction

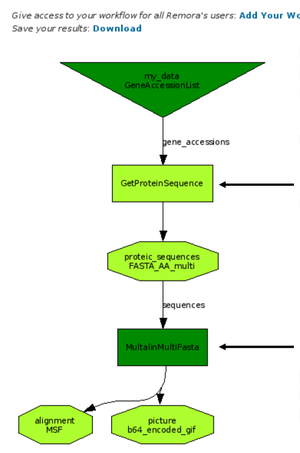




La construction du workflow se fait pas à pas: par simple click sur l'icône d'une sortie d'un service, le système propose l'ensemble des applications disponibles pour cette nouvelle donnée.

A tout moment de la conception de sa chaîne de traitement, l'utilisateur peut supprimer une branche entière par simple click sur le service à la racine.

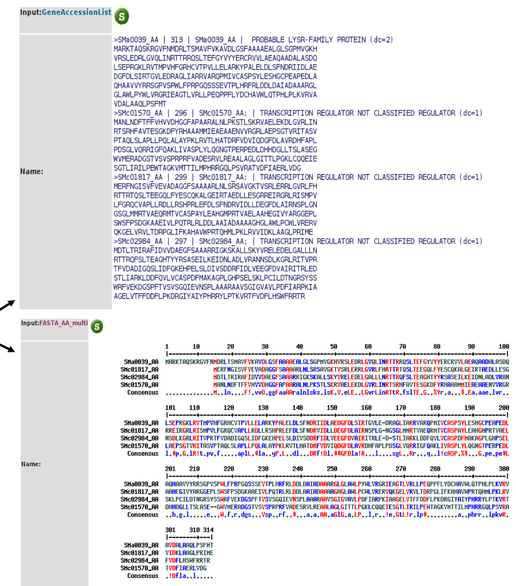
Résultats



L'utilisateur est prévenu par mail lorsque l'exécution du workflow est terminée (ou a échoué). L'accès aux résultats se fait via la représentation graphique:

1) en cliquant sur un service, les informations sur l'éxécution du service sont affichées: paramètres, description, exceptions levées.

2) en cliquant sur un noeud de sortie de service, la donnée produite est affichée avec un rendu **HTML** et un lien vers la donnée brute ainsi que vers la donnée ayant été utilisée en entrée du service.



Les résultats sont conservés trois semaines sur le serveur mais peuvent être sauvegardés sur le poste client (archive ZIP). Ainsi l'utilisateur peut naviguer dans ses résultats localement mais également rejouer son scénario d'analyse avec de nouvelles entrées et/ou de nouveaux paramètres via la section **Upload.**

Services

Mode Veille: REMORA propose l’exécution périodique du workflow à l’utilisateur. Cette fonctionnalité peut s’avérer importante par exemple dans le cadre de la recherche d’homologues dans une banque génomique d’un projet de séquençage.
Deux périodicité sont proposées: hebdomadaire ou mensuelle.

Section Frequently Asked Workflow (FAW): La conception et la configuration d’un workflow complexe pouvant s’avérer longue, REMORA propose aux utilisateurs de mettre à disposition de la communauté d’utilisateurs leurs chaînes de traitement.
Ces workflows sont alors décrits et déposés sur le serveur. Cette fonctionnalité va dans le sens du travail collaboratif à la base des web-services.

Mode Hook: Destiné aux développeurs, cette fonctionnalité permet l’exécution d’un web-service ou d’un workflow du FAW via une simple URL.

Bibiographie

[1] Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M.R., Wipat, A., and Li, P. (2004). Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* 20, 3045-3054.
[2] Carrere S. and Gouzy J. (2006). REMORA: a pilot in the ocean of BioMoby web-services *Bioinformatics* 22(7):900-1.
[3] Wilkinson, MD, Links, M. (2002). BioMOBY: an open-source biological web services proposal. Briefings In Bioinformatics 3:4. 331-341.

Formations

1. Développement de web-services BioMOBY en Perl (INRA, IPBS, Genopole)
2. Formation utilisateurs REMORA (LIPM)

Disponibilité

http://bioinfo.genopole-toulouse.prd.fr/remora