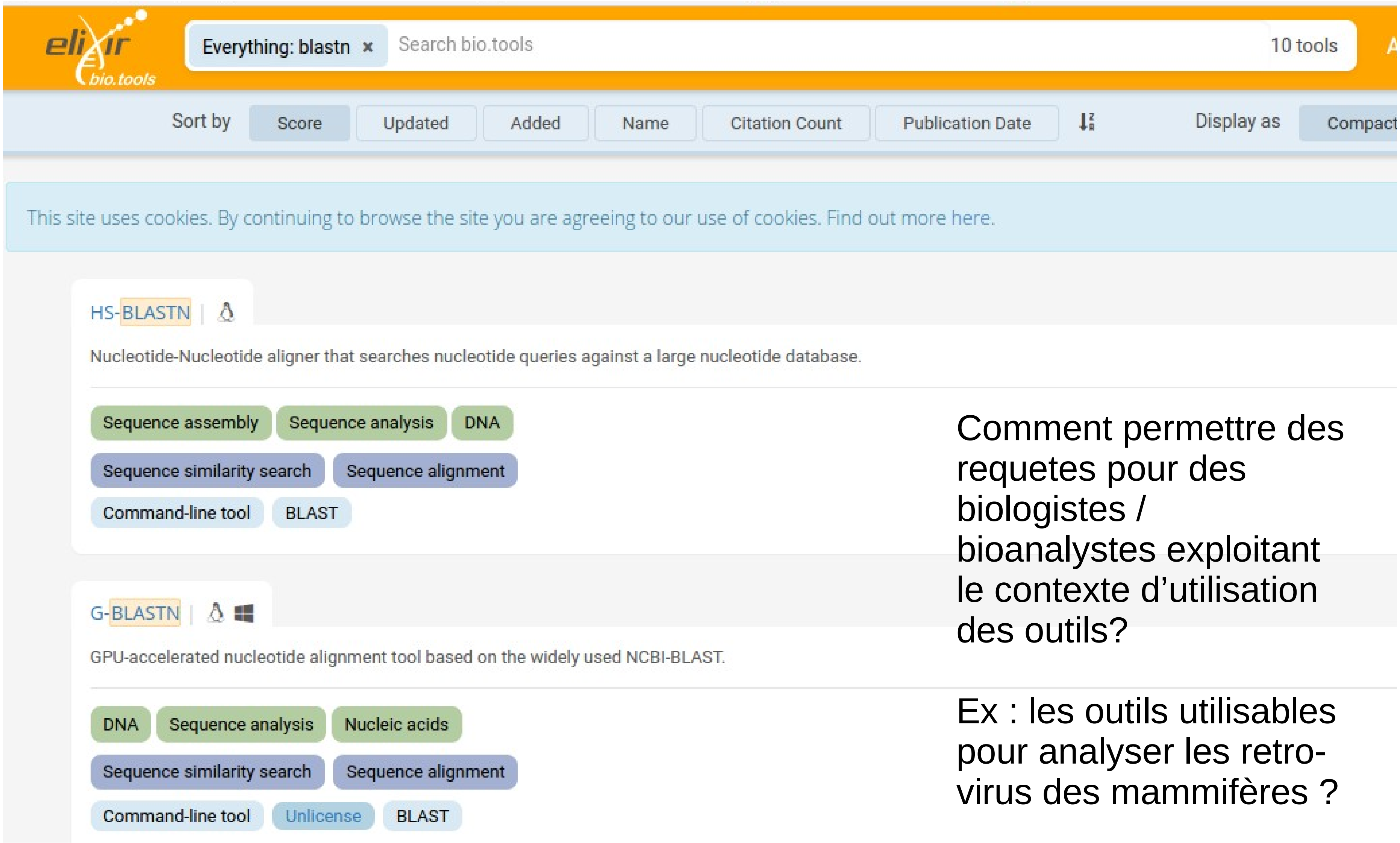
BIO4T : Contribution à la création d'un moteur de recherche de logiciels.

Application à l’analyse de données en biologie

Francois Moreews

fmoreews@irisa.fr

BIO4T : une contribution au moteur de recherche BIO.TOOLS


elixir
bio.tools
Sort by
Score
Updated
Added
Name
Citation Count
Publication Date
Display as
Compact

Everything: blastn × Search bio.tools 10 tools

This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here.

HS-BLASTN

Nucleotide-Nucleotide aligner that searches nucleotide queries against a large nucleotide database.

Sequence assembly | Sequence analysis | DNA
Sequence similarity search | Sequence alignment
Command-line tool | BLAST

G-BLASTN

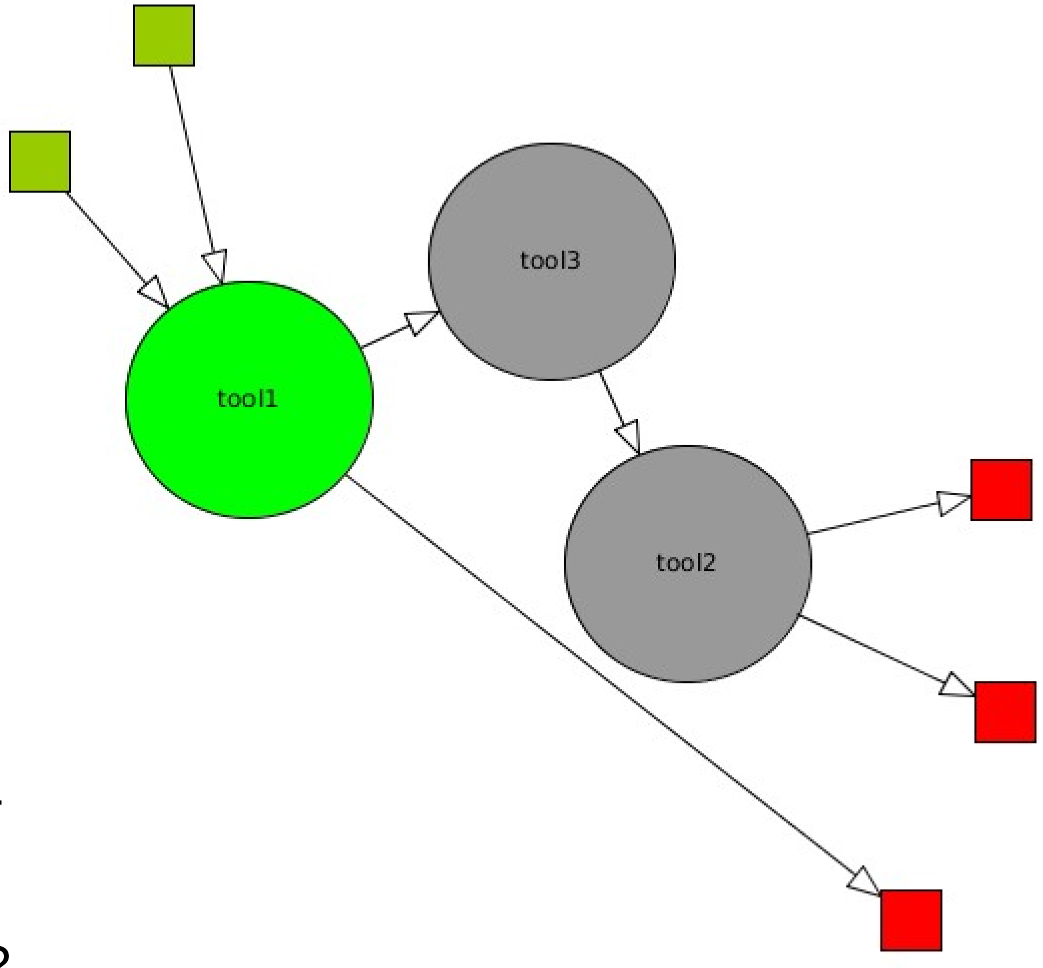
GPU-accelerated nucleotide alignment tool based on the widely used NCBI-BLAST.

DNA | Sequence analysis | Nucleic acids
Sequence similarity search | Sequence alignment
Command-line tool | Unlicense | BLAST

Comment permettre des requetes pour des biologistes / bioanalystes exploitant le contexte d’utilisation des outils?

Ex : les outils utilisables pour analyser les retro-virus des mammifères ?

BIO4T : proposer des workflows d’outils


tool3
tool1
tool2

Workflow / chaînes de traitement

Avec quel autres outil tool1 est il utilisé ?

Sur quel type de données ?

BIO4T : association des outils en workflow

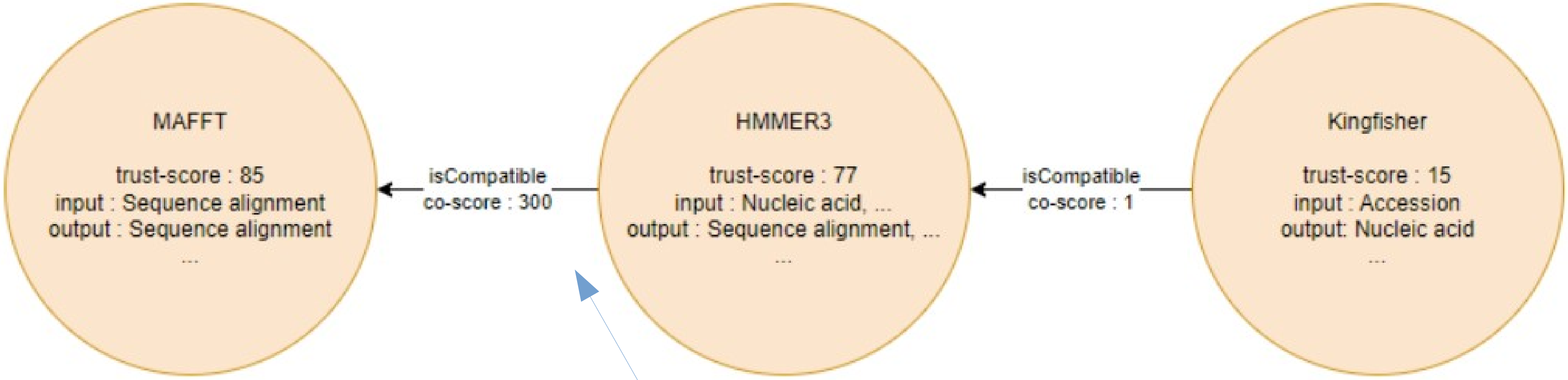

MAFFT
trust-score : 85
input : Sequence alignment
output : Sequence alignment
...
isCompatible
co-score : 300
HMMER3
trust-score : 77
input : Nucleic acid, ...
output : Sequence alignment, ...
...
isCompatible
co-score : 1
Kingfisher
trust-score : 15
input : Accession
output: Nucleic acid
...

Fig. 9 : Exemple de workflow du graph de requête

score de probabilité d’association entre noeuds du graphe =>

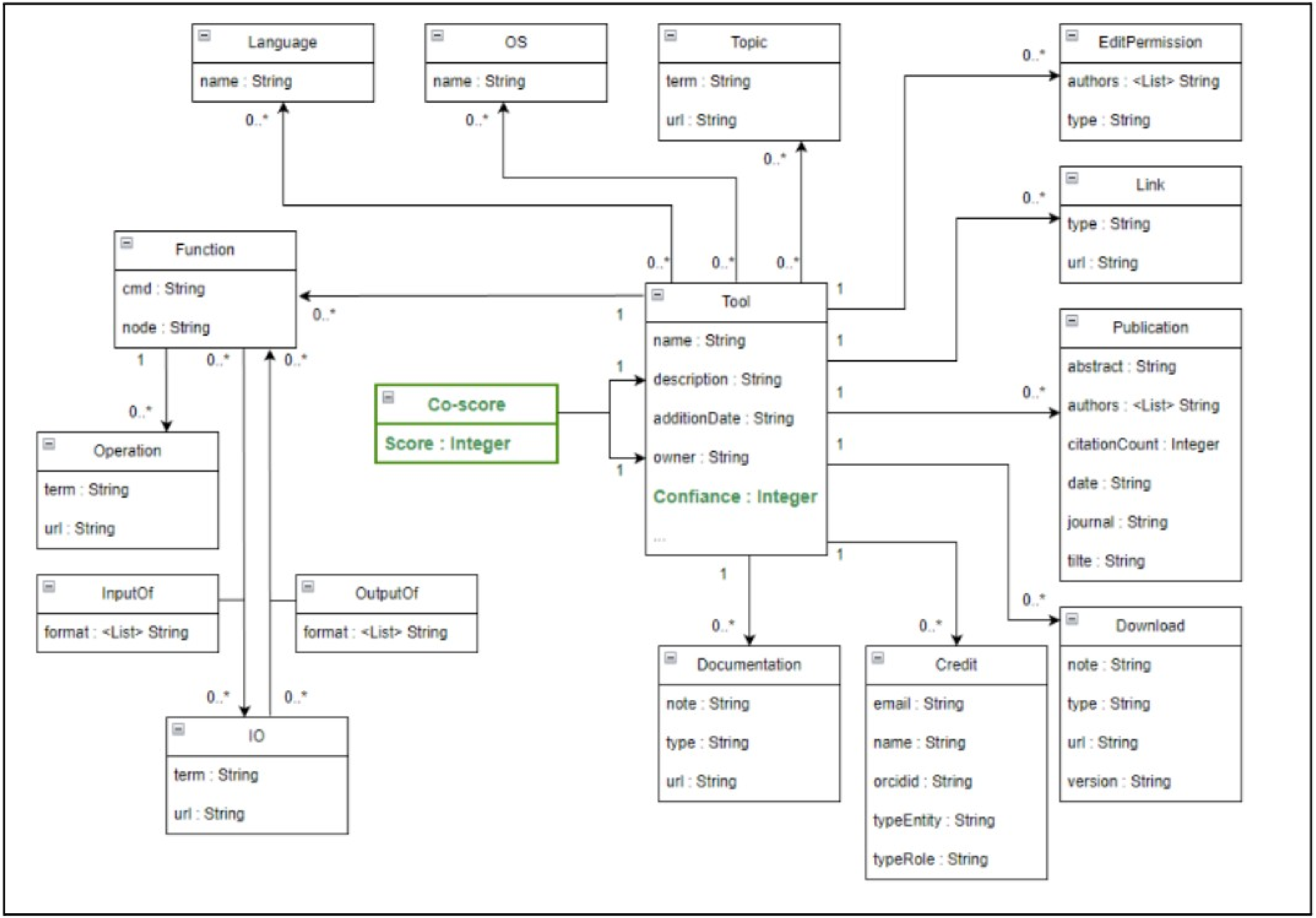
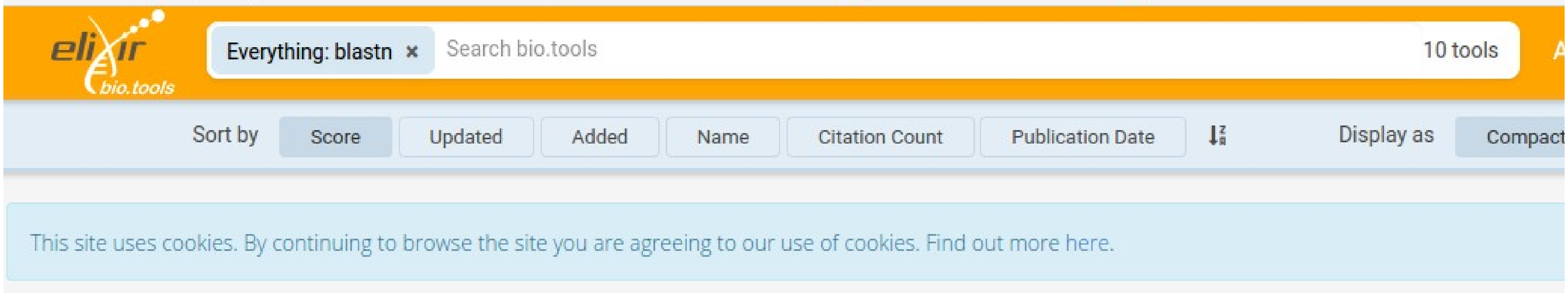

Language
name : String
0..*
OS
name : String
0..*
Topic
term : String
url : String
0..*
EditPermission
authors : <List> String
type : String
0..*
Link
type : String
url : String
0..*
Function
cmd : String
node : String
0..*
1
0..*
0..*
0..*
0..*
Tool
name : String
description : String
additionDate : String
owner : String
Confiance : Integer
...
1
1
1
1
1
1
1
Co-score
Score : Integer
Publication
abstract : String
authors : <List> String
citationCount : Integer
date : String
journal : String
tilte : String
0..*
0..*
Operation
term : String
url : String
InputOf
format : <List> String
OutputOf
format : <List> String
0..*
0..*
IO
term : String
url : String
0..*
Documentation
note : String
type : String
url : String
0..*
Credit
email : String
name : String
orcidid : String
typeEntity : String
typeRole : String
0..*
Download
note : String
type : String
url : String
version : String

Fig. 4 : Diagramme de classe de Bio4t

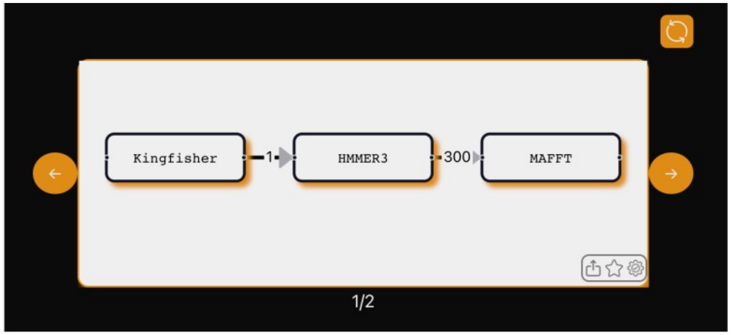
BIO4T : interface web

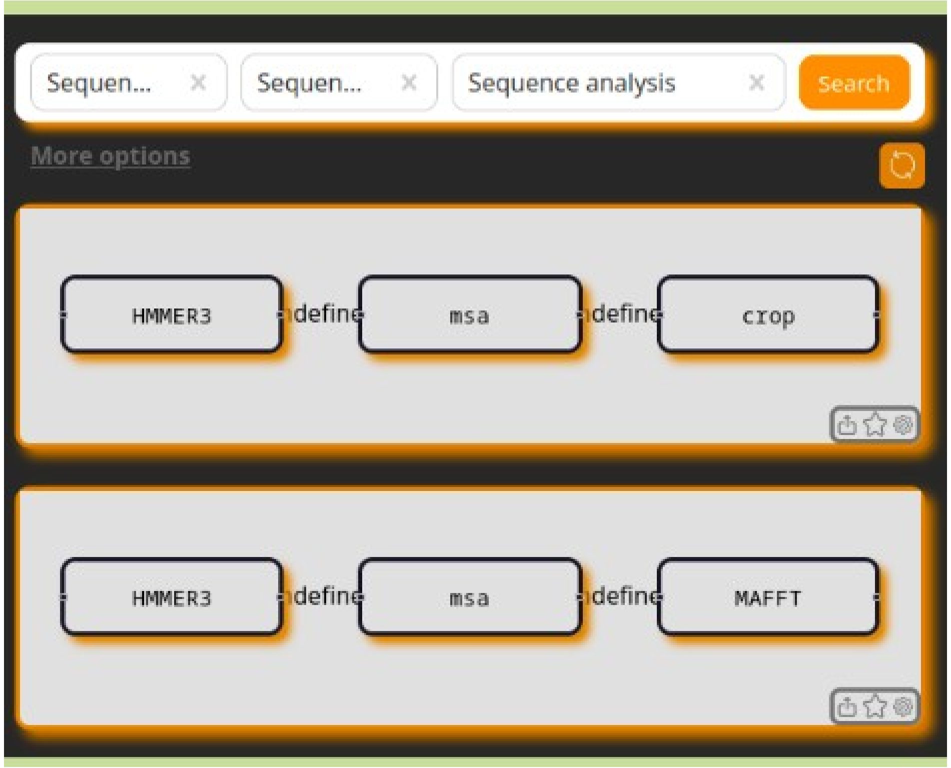

Everything: blastn ×

10 tools

This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here.

elixir
bio.tools
Search bio.tools
Sort by
Score
Updated
Added
Name
Citation Count
Publication Date
Display as
Compact


Kingfisher
1
HMMER3
300
MAFFT
1/2

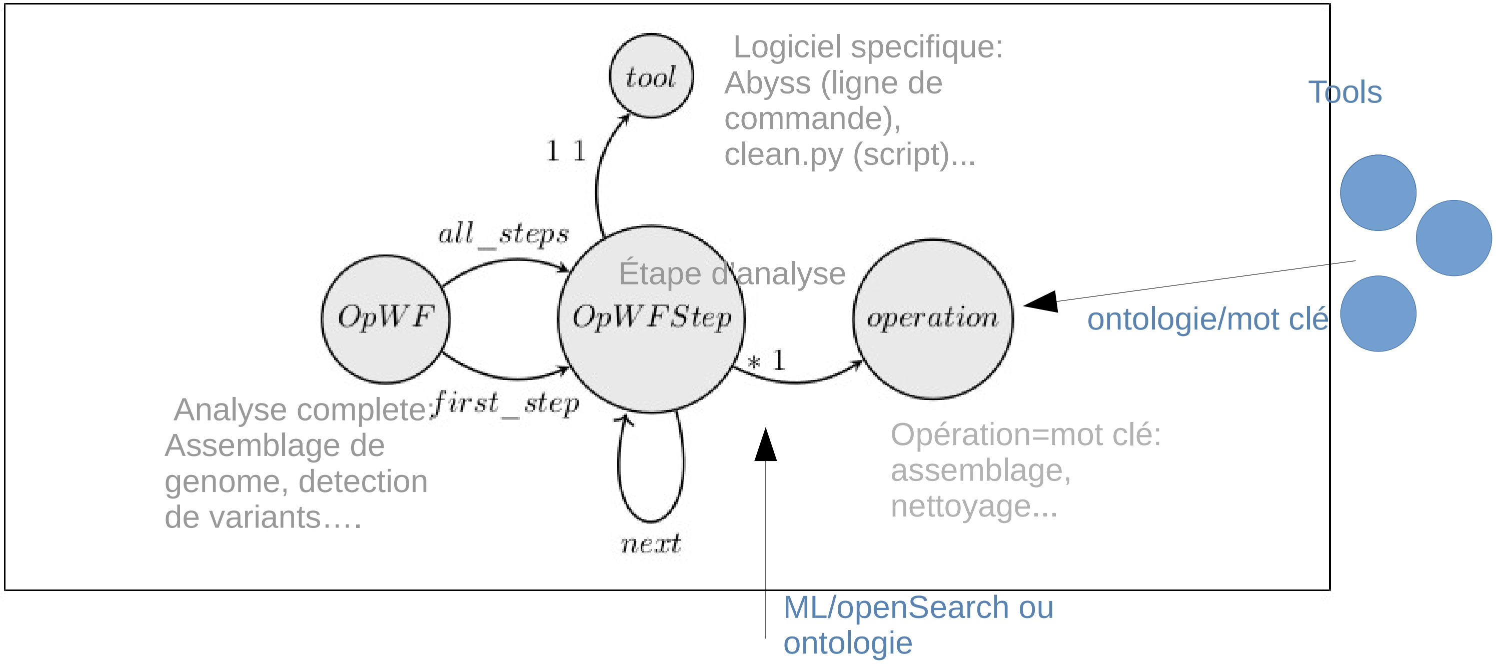

Sequen... ×

Sequen... ×

Sequence analysis ×

Search
More options
HMMER3
define
msa
define
crop
HMMER3
define
msa
define
MAFFT

Méthode Creation de workflows abstraits/théoriques à partir de données réelles


tool
1 1
all_steps
OpWF
OpWFStep
operation
∗ 1
first_step
next

Logiciel specifique:
Abyss (ligne de commande),
clean.py (script)...

Tools

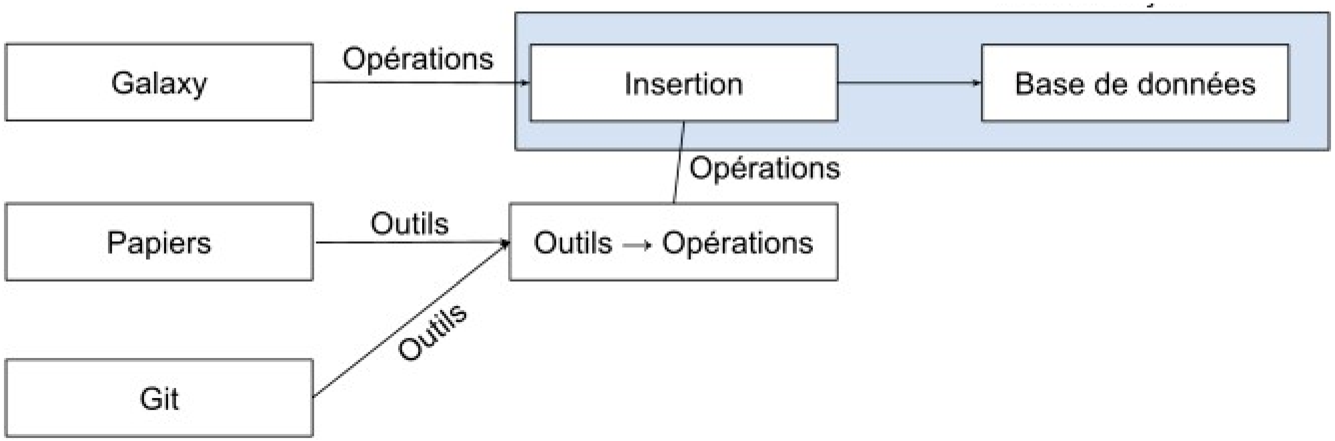
Étape d'analyse

ontologie/mot clé

Analyse complete:
Assemblage de genome, detection de variants….

Opération=mot clé:
assemblage, nettoyage...

ML/openSearch ou ontologie

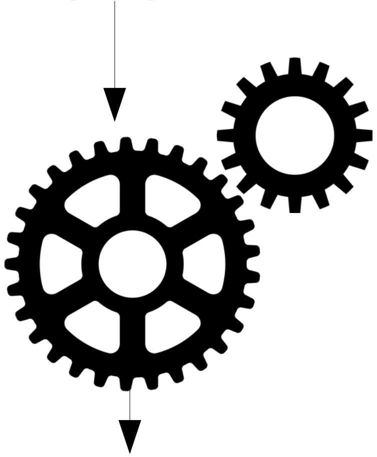

Galaxy
Opérations
Insertion
Base de données
Opérations
Papiers
Outils
Outils → Opérations
Outils
Git

Collecte des données

Modele en base de données

Objectif 1 : acquisition de données

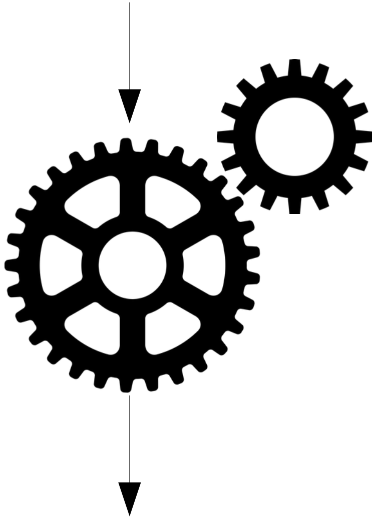
Workflows existants (Analyse automatique de la littérature scientifique, de resultat de recherche web, de git...) = description de l’execution d’une suite de logiciels ou scripts pour realiser uen tache scientifique precise



workflows “Modeles” ou abstrait= suite d’operation. Ex: 1 nettoyage des fragments d’adn→2 assemblage des fragments→ 3 annotation des sequences

Objectif 2 : text2workflow

Requete sous forme de text libre
“workflow for processing paired-end fastq files including quality control, adapter removal, and peak calling”



propositions concrètes de workflow d’analyse de données (liste d outils chainés)

BIO4T : Création d'un moteur de recherche de logiciels appliqué à l’analyse de données en biologie

TECHNOLOGIES

OpenSearch, ML,NLP,ontologies
PYTHON, API/WEB
GRAPHE/NEO4J

Francois Moreews

fmoreews@irisa.fr