

Projet international GBIF 2004-2008

Etat de situation au 1 février 2006

F. Burri, M. Bouzelboudjen

Service Informatique et télématique
Université de Neuchâtel

1. Introduction

L'Université de Neuchâtel réalise le noeud informatique de GBIF Suisse et met à disposition des chercheurs, étudiants et du public un réseau de compétences et de connaissances dans le domaine de la biodiversité. Ce noeud informatique a pour mandat d'offrir une plate-forme nationale assurant la fiabilité et la pérennité des flux d'informations. Dans ce cadre il développe et administre des bases de données accessibles via Internet dans les domaines de la faune principalement et de la flore (spécimens des collections scientifiques des Musées de Suisse).

2. Travaux réalisés durant l'année 2006

Le noeud informatique suisse a déployé d'importants efforts pour répondre au défi que représente l'intégration des données issues des différentes cultures linguistiques Suisse dans le réseau international GBIF. Ces efforts se sont traduits par le développement et la mise en œuvre d'outils logiciels à même de réaliser des opérations de validation et d'enrichissement des données. Ces outils logiciels ont été [présentés](#) lors de la journée d'information du 8 novembre 2006 à Berne. Ces développements ont été réalisés en tenant compte des exigences fixés par le comité scientifique de GBIF, à savoir:

- Réduire au maximum l'effort qui doit être fourni par les institutions pour publier leurs propres données sur le réseau international de GBIF
- Garantir un niveau de qualité des données conforme aux exigences fixées par le comité scientifique de GBIF Suisse

Pour répondre à ces exigences, la base de données du noeud informatique suisse a été construite sur la base d'un thésaurus hiérarchique multi-langue. Ce thésaurus porte actuellement sur 14 des 49 concepts prévus dans la base de données. Il offre quatre fonctionnalités essentielles à savoir :

1. Validation du contenu

Le processus de validation par le biais du thésaurus s'opère en comparant la valeur du champ fourni par l'institution avec les valeurs de référence contenues dans le thésaurus. D'autres règles de validations définies par le comité scientifique de GBIF Suisse sont aussi appliquées aux données.

2. Correction automatique

La correction automatique permet de corriger les fautes de frappes courantes telles que par exemple Genève → Genève.

3. Enrichissement

Le processus d'enrichissement s'applique aux données contenues dans des champs validés par le thésaurus et qui sont reliés entre eux par une relation hiérarchique. A partir d'une valeur de niveau hiérarchique supérieure, le logiciel va automatiquement déterminer les valeurs de niveau hiérarchique inférieur. La systématique avec ses quatre niveaux hiérarchiques définis dans le thésaurus (1 : Phylum, 2 : Class, 3 : Order, 4 : Family), la stratigraphie avec ses 3 niveaux hiérarchiques (1 : Période, 2 : Epoque, 3 : Etage) sont autant d'exemple de ce type de données.

4. Traduction automatique vers une langue ou sur une valeur de référence

Le processus de traduction automatique s'applique à toutes les données validées par le thésaurus. Il permet au système de publier des données sur Internet dans une langue de référence ou selon une désignation internationale reconnue.

En mai 2006, les premières données à caractère botanique (« Collection Data on the herbarium and living plants of the CJBG ») fournies par le « Conservatoire et Jardin botaniques de la Ville de Genève » ont été intégrées dans la base de données du nœud informatique Suisse et publiées sur le portail International de GBIF. La solution d'intégration a fait l'objet d'une [présentation](#) au comité scientifique de GBIF Suisse dans sa séance du 18 mai 2006.

En novembre 2006, les premières données à caractère zoologique fournies par le « Naturhistorisches Museum Bern » ont été intégrées et publiées sur Internet. La publication de ces données a nécessité le développement, sur le nœud informatique suisse, d'un portail Internet capable de formater et d'afficher des images.

1. Travaux prévus en 2007

Les travaux prévus en 2007 sont découpés en trois grands axes

1. La base de données du nœud informatique Suisse de GBIF est maintenant prête à accueillir les données à caractère botanique et zoologique. Aussi, il est prévu d'intégrer progressivement les données de toutes les institutions adhérentes aux concepts de GBIF Suisse.
2. Les collections concernant les observations de parasites font souvent références à des observations concernant les hôtes. Dans ce contexte, il est important que le lien entre parasite et hôte puisse être maintenu et exploité par les chercheurs. Un des objectifs de l'année 2007 sera de permettre d'intégrer ce type de données.
3. Les fonctionnalités d'interrogation offertes par le nœud international de GBIF ne permettent pas d'exploiter toute la richesse de l'information contenue dans le thésaurus multi-langue du nœud informatique Suisse. Par exemple, sur le portail international de GBIF, une interrogation de données basées sur une désignation allemande ne va rien retourner si la donnée est publiée par le nœud suisse en anglais (le nœud sui). Aussi, pour combler cette lacune et pour offrir

Neuchâtel, le 1 février 2006

NB :

Les 5 documents cités dans le texte sont disponibles à l'adresse : <http://www.gbif.ch/page6215.html>