

Cas d'usage de la fédération d'identité à un niveau international.

Point de vue d'un fournisseur de service dans le contexte des
humanités numérique

Benjamin Husson
Bertrand Gaiffe

ATILF

Cas d'usage
de la
fédération
d'identité à un
niveau
international.

Benjamin
Husson
Bertrand
Gaiffe

- projet européen d'infrastructure
- Common LAnguage and Resources Infrastructure
- nous avons participé à la phase “préparatoire”
- ... mais la France a décidé de ne pas intégrer la phase de construction

- mettre à disposition des chercheurs
 - des ressources (corpus annotés ou non, lexiques, etc.)
 - des outils pour travailler sur ces ressources.
- l'idée est que les chercheurs puissent travailler sur des ensembles de données provenant de plusieurs centres de ressources répartis à travers l'Europe.
- et même, éventuellement, qu'ils utilisent des outils sur ces collections de données.

Si Clarin réussit, le nombre de chercheurs concerné est la somme des chercheurs en linguistiques des pays participants !

Ce dont on rêve (par exemple)

Cas d'usage
de la
fédération
d'identité à un
niveau
international.

Benjamin
Husson
Bertrand
Gaiffe

- on cherche une expression dans des textes situés sur plusieurs serveurs
- si on clique sur une ligne, on obtient un contexte plus large. (de la part d'un des serveurs)

The screenshot displays the Ortolang web application interface. At the top, there are navigation tabs: 'Documents', 'Frequencies', 'Phrase Finder', and 'Contexts'. A search bar contains the keyword 'anxiety'. Below the search bar, a table lists search results. Each row includes a document ID (e.g., 233, 3), a snippet of text, a 'KEYWORD' column containing 'anxiety', and a 'CANDIDATE' column with the name 'Richardson'. The 'DELIVERY' column shows dates like '02-2007' or '01-2008'. The table is scrollable, and a mouse cursor is visible over the first row. Below the table, there is a section for 'Contexts' showing a larger snippet of text from the document, with the word 'anxiety' highlighted in blue. The context text discusses political and social issues, mentioning 'the world and my country', 'anxiety', and 'anxiety'.

- On voudrait (idéalement) ne s'identifier qu'une fois !

Le problème est que toutes les ressources ne sont pas dans le domaine public !

- dans la jungle des licences portant sur les ressources, quelques unes se dégagent :
 - utilisation par la “communauté enseignement recherche”.
Soit en termes de licence, soit en terme légal ;
 - libre pour l'enseignement
 - libre pour la recherche.
- et, surtout, “communauté enseignement recherche” + conditions plus restrictives !

Si un utilisateur se constitue un corpus de textes provenant de vingt centres européens différents,

- il est difficilement envisageable qu'il s'authentifie sur un centre chaque fois qu'il veut accéder à l'un des textes !
- il est également inconcevable que chaque centre gère la totalité des utilisateurs potentiels !

La seule solution réaliste (?) est de passer par un mécanisme de SSO

...à l'échelle européenne...

- est pour l'instant un vœu pieux !
- Mais il existe (existait) au moins une fédération de fédération d'identité : **Kalmar** la fédération regroupant la Finlande, la Norvège, le Danemark, la Suède et l'Islande.
- L'idée a donc été (à l'initiative du MPI de Nimègue) de créer une "fédération de fédération" durant la phase préparatoire de Clarin.
- C'est dans ce cadre que nous avons travaillé.

- est pour l'instant un vœu pieux !
- Mais il existe (existait) au moins une fédération de fédération d'identité : **Kalmar** la fédération regroupant la Finlande, la Norvège, le Danemark, la Suède et l'Islande.
- L'idée a donc été (à l'initiative du MPI de Nimègue) de créer une "fédération de fédération" durant la phase préparatoire de Clarin.
- C'est dans ce cadre que nous avons travaillé.

Rq : EduGAIN n'était pas une option pour nous ; question de timing...

- de la phase préparatoire d'un projet !
- il ne s'agit même pas (à mon avis) de tests de faisabilité
- identifier les problèmes
- essayer de s'approprier des techniques
- viser à long termes (la phase de construction est prévue sur 5 ans). La phase d'exploitation doit (en théorie) commencer en 2016.

- minimaux :
 - eduPersonPrincipalName ou eduPersonTargetedId
- souhaités :
 - cn (common Name)
 - mail
 - o (organizationname) ou schacHomeOrganization

cf : <http://www.clarin.eu/node/3452>

- Nous souhaitons protéger certaines des ressources mises en lignes sur le site du CNRTL, fournisseur de service dans la fédération.
- Les restrictions d'accès à ces ressources sont variables et peuvent dépendre :
 - d'informations disponibles dans un profil SAML
 - d'informations liées spécifiquement à notre service
- Par exemple certaines ressources pourraient être accessibles à tout les étudiants ET aux personnes ayant souscrit à un abonnement à notre ressource.

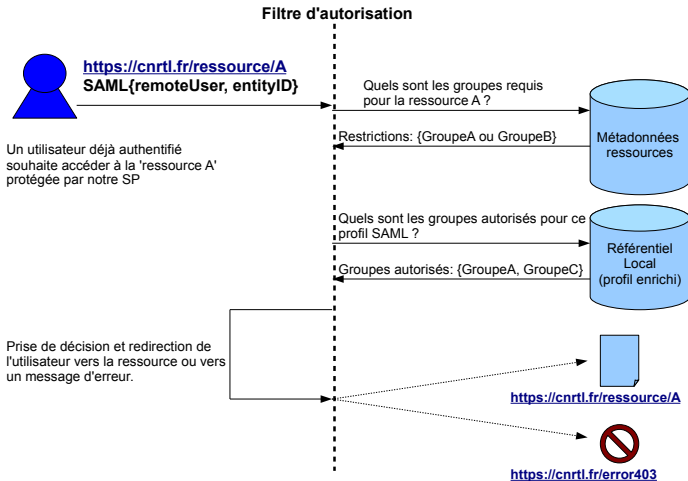
- Nous schématisons donc des groupes d'utilisateurs (enseignant-chercheur, abonnement_XXX, clarin, etc...) représentant les différents profils dans notre application
- A ces groupes nous associons des règles permettant d'y verser un utilisateur identifié à partir de son profil SAML.
- Une base locale nous permet d'enrichir le profil d'un utilisateur ou d'une entité en y ajoutant les attributs spécifiques à notre service.

- Pour écrire les règles nous avons commencé par choisir les attributs dont nous aurons besoin (~ 5) à partir du catalogue supann09 (qui contient en tout ~ 50 champs)
- Cependant les fournisseurs d'identité n'ont pas d'obligation quant à la valuation de ces champs.
- Il n'est donc pas facile de se fier à ces attributs pour gérer une autorisation fine, surtout dans le cadre d'un projet faisant intervenir virtuellement l'ensemble des IDP

Cas d'usage
de la
fédération
d'identité à un
niveau
international.

Benjamin
Husson
Bertrand
Gaiffe

- Les deux champs systématiquement présents sont les identifiants unique de l'utilisateur et de son fournisseur d'identité.
- Ces deux champs sont donc utilisés dans un référentiel local en tant qu'index d'un profil que nous aurons enrichi.



- A l'heure actuelle, la France n'est plus dans Clarin. En revanche, la France participe à Dariah. (Digital Research Infrastructure for the Arts and Humanities)
- Les mêmes problèmes se posent (poserons) dans le contexte "Art and Humanities".

Ortolang est un “Equipex” en réseau auquel participent :

- l'ATILF (Nancy)
- le LPL (Aix/Marseille)
- Modyco (Paris 10)
- le LLL (Orléans)

Nous ambitionnons de continuer à participer à Clarin (en tant que nœud) et à Dariah.

Ortolang est un “Equipex” en réseau auquel participent :

- l'ATILF (Nancy)
- le LPL (Aix/Marseille)
- Modyco (Paris 10)
- le LLL (Orléans)

Nous ambitionnons de continuer à participer à Clarin (en tant que nœud) et à Dariah.

Et nous avons plus de questions que de réponses !