



68th IFLA Council and General Conference

August 18-24, 2002

Code Number: 090-163-F
Division Number: I
Professional Group: National Libraries
Joint Meeting with: Information Technology
Meeting Number: 163
Simultaneous: -
Interpretation:

Accès aux Archives du Web : le projet d'accès aux archives nordiques du Web

Svein Arne Brygfjeld

Bibliothèque nationale de Norvège
Oslo, Norvège

Résumé : Les bibliothèques nationales des cinq pays nordiques ont lancé un projet commun concernant les principes, les méthodes et les outils d'accès aux archives de documents diffusés sur le Web. Ce projet a conduit à la réalisation d'un prototype de système d'accès grâce auquel un utilisateur peut feuilleter, naviguer et chercher à travers le temps et l'espace. Ce système d'accès est indépendant de l'archive elle-même et peut facilement s'adapter à un nouvel environnement. Moteurs de recherche et interfaces-utilisateur (basés sur Java) sont essentiels aux fonctionnalités désirées par les usagers.

1 Introduction :

L'archivage, ainsi que la conservation à longue échéance des documents publiés sur le Web (WWW), représentent un défi pour la communauté internationale chargée de la conservation de l'information. La collecte des données diffusées sur le Web, l'accès aux archives

du Web, sont des secteurs liés au domaine de la conservation à long terme. Les bibliothèques nationales des cinq pays nordiques (Danemark, Finlande, Islande, Norvège et Suède) se sont fortement impliquées dans la découverte de solutions pour collecter, archiver, et accéder aux archives.

Ces bibliothèques se sont réunies afin d'étudier les techniques et les méthodes consacrées à ces secteurs, une initiative intitulée « *Les archives Web Nordiques* ». Ces derniers dix-huit mois, les efforts ont porté majoritairement sur l'exploration des divers moyens d'accès aux archives Web. Nordunet2 a soutenu ce travail, permettant ainsi la mise en oeuvre d'un projet ciblé appelé NWA: Projet d'Accès aux Archives Nordiques du Web.

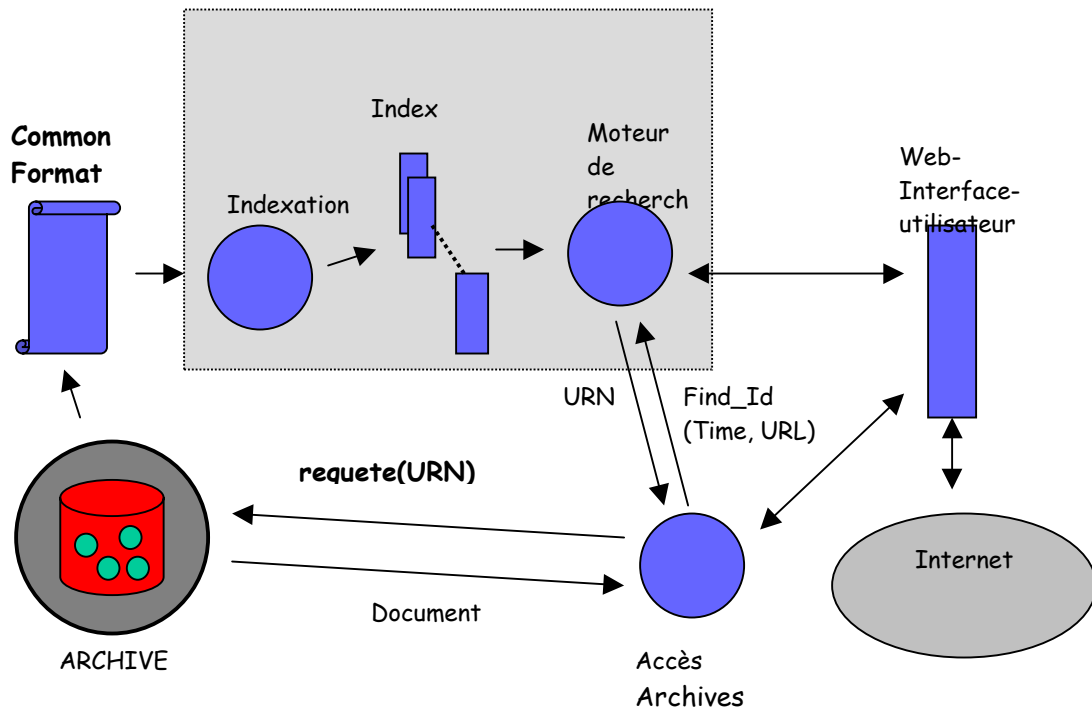
2. Objectifs :

L'objectif majeur du projet NWA est la création d'un système d'accès pour un archivage Web à grande échelle. Ce système offrira des méthodes d'accès bien connues telles que recherche, navigation et feuilletage. Il serait également souhaitable de pouvoir feuilleter et naviguer à travers le temps et l'espace. Le projet NWA repose sur l'hypothèse selon laquelle il existe plusieurs types d'archives, et que chaque type d'archives contient éventuellement plusieurs versions d'un nombre significatif de documents Web. De telles archives seront vraisemblablement construites de manières différentes; aussi, conception et mise en oeuvre du système d'accès devront-elles être indépendantes de la structure et de l'architecture interne des archives.

3. Choix architecturaux

En commençant à travailler sur l'architecture du système, il s'avéra souhaitable de le diviser en modules significatifs et bien délimités. Ainsi, on permettrait une meilleure distribution des travaux de développement à réaliser en collaboration. Les différents modules identifiés sont listés ci-dessous et présentés dans la figure 1 :

- Introduction du document : le système d'accès recevra des documents d'archive XML pour indexation. Ces documents contiendront des métadonnées de format « Dublin Core » prévoyant un support pour de l'OAI. Il y aura aussi des métadonnées d'archivage, par exemple la date de la collecte.
- Indexation: organisation des documents en perspective des requêtes.
- Moteur de recherche: élément destiné à soutenir la recherche parmi les documents indexés et les métadonnées.
- Interface Web : élément destiné à soutenir l'interface utilisateur reposant sur l'usage du World Wide Web
- Accès à l'archive : élément destiné à soutenir la fourniture du document depuis l'archive web qui lui correspond.



4 . Stratégies et solutions de mise en œuvre

La vitesse du service doit être aussi performante que celle des moteurs de recherche sur internet. D'où la nécessité d'utiliser un moteur de recherche très compétitif pour servir recherche, navigation et feuilletage. L'introduction d'un tel élément de logiciel a simplifié notablement le développement, tout en produisant les performances attendues.

Par ailleurs, l'exécution est hautement modularisée. Chaque module est relativement petit, rendant possible une re-exécution si nécessaire. Cette approche permet de modifier également les fonctionnalités liées à l'introduction directe de nouveaux modules. Un gros effort a porté sur la réalisation d'une interface propre et clairement définie face au moteur de recherche. Le portage vers d'autres moteurs de recherche représenterait donc une autre tâche à accomplir.

5. Interface archive :

Comme nous l'avons montré précédemment, l'accès aux archives Web doit reposer sur les mêmes technologies que l'accès au Web. Un outil de feuilletage du Web qui fournirait la même empreinte du contenu que celle produite par l'original serait optimal.

6. Fonctionnalités

L'interface utilisateur devrait garantir les fonctions bien connues de recherche, navigation et feuilletage que les utilisateurs croisent chaque jour sur Internet. Il devrait aussi garantir ces fonctions sur les différentes versions des documents,

introduisant ce faisant le concept de «recherche, feuilletage et navigation à travers le temps et l'espace ».

6.1 Recherche :

La fonction recherche au sein de l'archive repose sur l'utilisation d'un moteur de recherche qui lui est dédié. L'intérêt d'introduire un tel moteur de recherche, c'est la capacité d'offrir une vitesse satisfaisante à l'utilisateur. La recherche doit fonctionner à la fois sur les métadonnées ainsi que sur le contenu des documents. Un des défis à relever, tout en respectant la présentation des résultats, serait de pouvoir atteindre avec succès plusieurs versions de nombreux documents élargissant ainsi le problème des grandes séries de résultats.

6.2. Feuilletage

Comme sur le Web, l'utilisateur devrait pouvoir feuilleter l'archive simplement, en suivant les liens à travers les documents qui s'affichent. L'utilisateur devrait aussi pouvoir feuilleter les différentes versions d'un document donné en utilisant l'axe-temps , basé sur java, et présenté dans la figure 2. En cliquant sur les flèches, on obtiendrait la version précédente ou la version suivante du document d'archive.

6.3 Navigation :

En utilisant des indicateurs de localisation connus (URL) l'utilisateur devrait pouvoir se situer et naviguer dans l'archive. Ainsi, pour une date précise, l'archive devrait fournir un document quand un utilisateur tape l'URL original de ce même document. Inversement, étant donné un URL original, l'archive devrait pouvoir fournir un document relié à une date précise quand l'utilisateur donne une date précise. Ces fonctions sont montrées à travers les champs *lieu* et *temps* de la figure 3



Figure 2.



Figure 3.

5 Liens, bibliographie :

- 1 Bibliothèque royale du Danemark, <http://www.kb.dk>
- 2 Bibliothèque nationale et universitaire, Helsinki, Finlande,
- 3 Bibliothèque nationale et universitaire d'Islande,
- 4 Bibliothèque nationale de Norvège, <http://www.nb.no>
- 5 Bibliothèque royale de Suède, <http://www.kb.se>
- 6 Archives Web, nordiques, <http://www.nwa.nb.no>
- 7 Nordunet2, <http://www.nordunet2.org>
- 8 Kulturarw3, Bibliothèque royale de Suède, <http://kulturarw.kb.se/html/kulturarw3.eng.html>
- 9 The Wayback Machine, <http://www.archive.org>
- 10 Zeitschrift für Bibliothekswesen und Bibliographie, n° ¾ 2001
- 11 NEDLIB project, <http://www.kb.nl/coop/nedlib>
- 12 OAIS *Modèle de référence pour un système d'information d'archives ouvertes*,
<http://ssdoo.gsfc.nasa.gov/nost/isoas/>
- 13 Extensible Markup Language, <http://www.w3.org>
- 14 Initiative Archives ouvertes, <http://www.openarchives.org>
- 15 Recherche rapide et transfer, <http://www.fast.no>
- 16 Initiative Sources ouvertes, <http://www.opensource.org>
- 17 RLG DigiNews 2001, volume 5, numero 2, <http://www.rlg.org/preserv/diginews/diginews5-2.htm>

Traduction / Christiane BARYLA
Ecole française de Rome