

Formation à la recherche documentaire sur le web

Objectif :

- Savoir effectuer des recherches pertinentes sur Internet
- Savoir distinguer les informations fiables sur Internet
- Savoir effectuer des recherches dans des ressources en ligne de bibliothèque

La recherche sur Internet

Internet est un réservoir tellement gigantesque d'informations qu'il convient d'en connaître les mécanismes pour faire des recherches efficaces.

Différents outils permettent d'accéder au web :

Moteurs	outil automatisé de recherche de ressources du web	Google, Bing, Exalead, Altavista
Métagoteurs	outil de recherche qui interroge en parallèle plusieurs outils (moteurs / annuaires)	Ixquick,
Annuaires	répertoire de sites dont la qualité, la pertinence et la fiabilité ont été évaluées	Dmoz

1. Les moteurs de recherche

Avantages des moteurs de recherche :

- La possibilité de faire des recherches dans une grande masse d'informations.
- Le fait d'obtenir rapidement des informations précises sur des sujets divers et variés.
- Le fait de pouvoir trouver des documents spécifiques.
- Le fait de repérer des sites récents ou ayant été écartés des annuaires.
- Le fait de pouvoir réaliser des recherches complexes en utilisant la logique booléenne.

Inconvénients des moteurs de recherche :

- Absence de contrôle des informations (présence d'URL périmées dans les résultats).
- Des interrogations qui semblent complexes car les interfaces changent d'un moteur de recherche à l'autre.
- Des résultats parfois surprenants car l'indexation est automatique.
- Des résultats qui sont classés selon un ordre qui donne la priorité à la popularité des informations
- Attention : les moteurs n'ont pas accès au web invisible (tout ce qui se trouve derrière un formulaire)

a. Syntaxe des principaux moteurs

La plupart des outils de recherche sur internet utilisent la syntaxe suivante :

- Rechercher l'expression exacte : utiliser les guillemets « »
- Imposer un terme : mettre un + (collé à gauche du terme concerné)
- Exclure un terme : mettre un - (collé à gauche du terme concerné)
- Faire une recherche sur au moins un des mots : utiliser OU/OR (utiliser les majuscules)

Les moteurs de recherche peuvent tenir compte des accents.

Exercice : Cherchez électricité puis electricite dans google

Une plus grande importance est donnée au premier mot choisi.

Exercice : Cherchez paris dakar puis dakar paris dans google.

Le nombre de résultats est en général si important (même si sur la dernière page le nombre de résultats effectif n'est pas celui initialement annoncé par google) qu'il est intéressant de limiter la recherche par d'autres moyens. Voici quelques outils qui permettent de restreindre la recherche dans google :

Pour limiter la requête au titre des pages : *intitle*

Exercice : Recherchez « crise financière » puis « intitle :crise financière»

Pour limiter la requête sur Google à l'url : *inurl*

Pour limiter la recherche à un site ou à un domaine spécifique : *site : www....*

Regardez le formulaire de recherche avancée sur google : d'autres filtres (langue etc) sont proposés.

Le site Internet Abondance propose des fiches pour chaque outil de recherche, présentant les spécificités de syntaxe de chacun. <http://outils.abondance.com/>

b. Connaissez-vous google scholar ? <http://scholar.google.fr/>

Ce moteur de recherche spécialisé dans la littérature de recherche porte sur des travaux universitaires.

Exercice : Faites une recherche sur la crise financière. Quelles différences relevez-vous par rapport à votre recherche dans google ?

Exercices (*utilisation des formulaires de recherche avancée dans des moteurs de recherche*) :

- a. Cherchez des articles scientifiques écrits par Alan Turing dans Google scholar.
- b. Cherchez des informations rédigées en anglais sur Alan Turing en utilisant Bing.
- c. Cherchez des documents PDF en français sur Alan Turing en utilisant Google.
- d. Cherchez une biographie de Alan Turing sur des sites hébergés au Royaume-Uni en utilisant Yahoo.
- e. Cherchez une biographie rédigée en français de Alan Turing en utilisant Exalead.

2. **Les métamoteurs** (Par exemple Ixquick <http://www.ixquick.com>)

Ixquick transforme la question posée dans la syntaxe de chaque outil interrogé. Ixquick élimine les doublons et donne les « top 10 » résultant de chaque outil utilisé.

3. **Les annuaires**

Il s'agit d'un classement sélectif et raisonné de sites web selon un thème donné. Construits « à la main », après sélection plus ou moins sommaire, ils sont constitués de rubriques (ou catégories), elles-mêmes divisées en sous-rubriques.

Avantages :

- Trouver des sites dans un domaine donné
- Explorer un sujet

Inconvénients :

- Moins d'exhaustivité
- Mise à jour moins rapide

a. **Exemple : Open Directory**

<http://www.dmoz.org/>

L'Open Directory Project (OPD), aussi appelé DMOZ (Directory Mozilla) est un répertoire de sites web créé en 1998. Il est géré, développé et maintenu par des éditeurs bénévoles, chacun de ceux-ci étant responsable de la vérification des sites et de leur classement. DMOZ est un projet gratuit pour l'utilisateur et collaboratif.

b. **Exemple : signets des universités**

<http://www.signets-universites.fr/>

La fiabilité des informations sur Internet

Internet est un formidable réservoir d'informations mais attention :

Tout le monde peut publier de l'information en ligne: certaines informations sont de haute qualité, d'autres laissent à désirer...

Il est difficile de contrôler ce qui est publié: pas toujours de responsabilité éditoriale globale, peu ou pas de filtre...

Les informations peuvent changer ou être modifiées à tout moment.

Il faut donc se montrer vigilant et développer son esprit critique face à l'absence de contrôle éditorial sur internet, l'abondance de publicités, la confusion possible entre les liens commerciaux et les résultats, etc.

Évaluer l'information, reconnaître une information fiable sont des opérations difficiles mais voici quelques outils utiles et quelques questions à se poser.

Démarche, questions à se poser :

1. QUI est la source ? : évaluation de la source ou de l'auteur

Est-il clairement identifié? Quelles sont ses compétences, ses fonctions? Est-il connu? L'auteur apparaît-il lui-même dans une bibliographie?

Peut-on entrer en contact avec lui? (adresse, téléphone, e-mail, adresse d'une page personnelle...)

Est-ce un organisme institutionnel ou pas? S'il s'agit d'une organisation, est-elle connue? Y a-t-il moyen de vérifier le sérieux de l'organisation?

2. De QUOI est-il question ?

Quel est le sujet ou thème général annoncé ?

Est-il possible de vérifier l'information en consultant d'autres sources, en ligne ou non ?

L'information fait-elle le tour du sujet ? Donne-t-on plus d'un point de vue ?

Y a-t-il des liens vers d'autres sites qui fournissent des points de vue différents ? Le site réfère-t-il à d'autres sources d'information plus complètes ou offre-t-il des liens vers ces sources ?

3. COMMENT l'information est-elle présentée ?

Est-ce que certaines informations sont manquantes ?

L'information est-elle facile à comprendre ?

L'information est-elle cohérente ?

Les liens présentés sont-ils intéressants ?

4. De QUAND date l'information ?

Quelle est la date de création du site?

Quelle est celle de sa dernière mise à jour? Cette date est-elle affichée dans la page?

L'information mise à disposition est-elle à jour?

Les liens vers d'autres sites sont-ils à jour?

Attention : de nombreux sites sont abandonnés après leur création = l'information présentée risque d'être périmée...

5. OÙ est située l'information ?

L'URL est-elle explicite ?

Par quel serveur le site est-il hébergé ?

Connaît-on la nationalité de l'auteur ou de l'hébergeur ?

Les noms de domaine sont en eux-mêmes une indication.

Exercice: Dans la liste suivante, indiquez à quoi correspondent les extensions dans les URL:

<http://pageperso.lif.univ-mrs.fr/~bertrand.estellon/>

<http://www.education.gouv.fr/>

<http://www.enseignement.be/>

<http://www.ogm.org/>

<http://ogm.gouv.fr/>

Rappel sur les noms de domaine:

– génériques = par rapport à l'activité de l'organisme.

- .org pour organisme à but non lucratif
- .com pour organisme à but lucratif
- .gouv (ou .gov) pour les organismes gouvernementaux

– géographiques :

- .fr pour France
- .ch pour la Suisse

6. L'ergonomie

La navigation se fait-elle facilement sans être perdu?

Est-il tjrs possible de dire où l'on se situe dans le site?

Le site est-il «en construction »?

Les pages ont-elles toutes un titre explicite?

Existe-t-il une aide en ligne?

Le site possède-t-il un moteur de recherche interne?

Les ressources sont-elles bien organisées et présentées logiquement?

Attention : un site ou un document du web peut être parfaitement organisé et très beau au plan visuel mais avec un contenu erroné. L'inverse est possible : un document dont le contenu est pertinent et l'auteur fiable peut être mal structuré, avec une consultation difficile.

Ressources en ligne de la bibliothèque

NB : Deux idées erronées sont très répandues concernant le Web:

- Tous les sites sont à accès gratuit. *Au contraire*, la plupart des meilleures banques de données requièrent un abonnement et ne sont pas gratuites.
- L'objet de notre recherche se trouvera sur le site lui-même. *Au contraire*, vous ne trouverez souvent que des références bibliographiques aux documents, tout au plus un résumé, rarement le texte intégral du document.

Les informations disponibles gratuitement sur Internet sont rédigées par des producteurs de sites Web qui travaillent en conscience (ou pas !) et sous leur responsabilité ... mais sans garantie particulière vis à vis de leurs lecteurs. La validité de l'information proposée n'est donc aucunement garantie. Par contre, quand les services sont payants, la responsabilité du producteur ou de l'éditeur est engagée.

La bibliothèque est abonnée à des ressources en ligne (payantes) qui sont fiables. Comme dans le cas de publications papier, la validation (comité éditorial) se fait avant la publication en ligne.

<http://bu.univmed.fr/>

1. Ressources électroniques (sous l'onglet « ressources »)

The screenshot shows the website of the University of the Mediterranean's library. The browser window title is "Bibliothèque de l'Université de la Méditerranée - Mozilla Firefox". The address bar shows the URL "http://bu.univmed.fr/flora/servlet/LoginServlet". The page header includes the university logo and the text "UNIVERSITÉ DE LA MÉDITERRANÉE AIX-MARSEILLE II" and "SERVICE COMMUN DE LA DOCUMENTATION". A navigation menu contains "Accueil", "Recherche", "Ressources", "Bibliothèques", "Services", and "S'authentifier". The "Ressources" section is active, displaying a sidebar menu with options like "Notre catalogue", "Autres catalogues", "Livres électroniques", "Revue électronique", "Bases de données", "Annale d'examens", "Sélection de sites web", "Thèses électroniques", and "Archives ouvertes". The main content area features a grid of resource categories, each with a small image and a square icon: "Catalogue de la Bibliothèque", "Livres électroniques", "Revue électronique", "Bases de données scientifiques", "Annales d'examens", "Thèses électroniques de l'université", "Archives ouvertes", and "Sélection de sites web". The Windows taskbar at the bottom shows the system tray with the time "13:10" and the language "FR".

a. Livres électroniques

i. SAFARI : 101 titres d'informatique (en anglais)

Sélectionnez un ouvrage que vous pouvez consulter en ligne

ii. Numilog

Sélectionnez un ouvrage que vous pouvez consulter en ligne

iii. Dictionnaires Larousse

Ce dictionnaire est un ouvrage de référence disponible en ligne.

Comparez Wikipédia (en ligne) / Encyclopédie Larousse (en ligne)

Faire par exemple la recherche « crise financière »

NB : dans l'encyclopédie Larousse, un double-clic sur un mot vous donne sa définition

Wikipédia	Larousse
anonyme	sourcé
collaboratif	Collaboratif + comité éditorial
Pas de contrôle systématique	Contrôle systématique
Contrôle des informations a posteriori	Contrôle des informations a posteriori ou a priori en fonction de l'auteur de l'article
Libre de droit	propriétaire
Hétérogénéité des contenus, qui varient fortement d'un article à un autre	Homogénéité des contenus (style, types d'information proposés, longueur)
Poids de l'actualité	Moindre poids de l'actualité
Contenus mouvants	Contenus évolutifs cf contributions (seul l'auteur modifie)

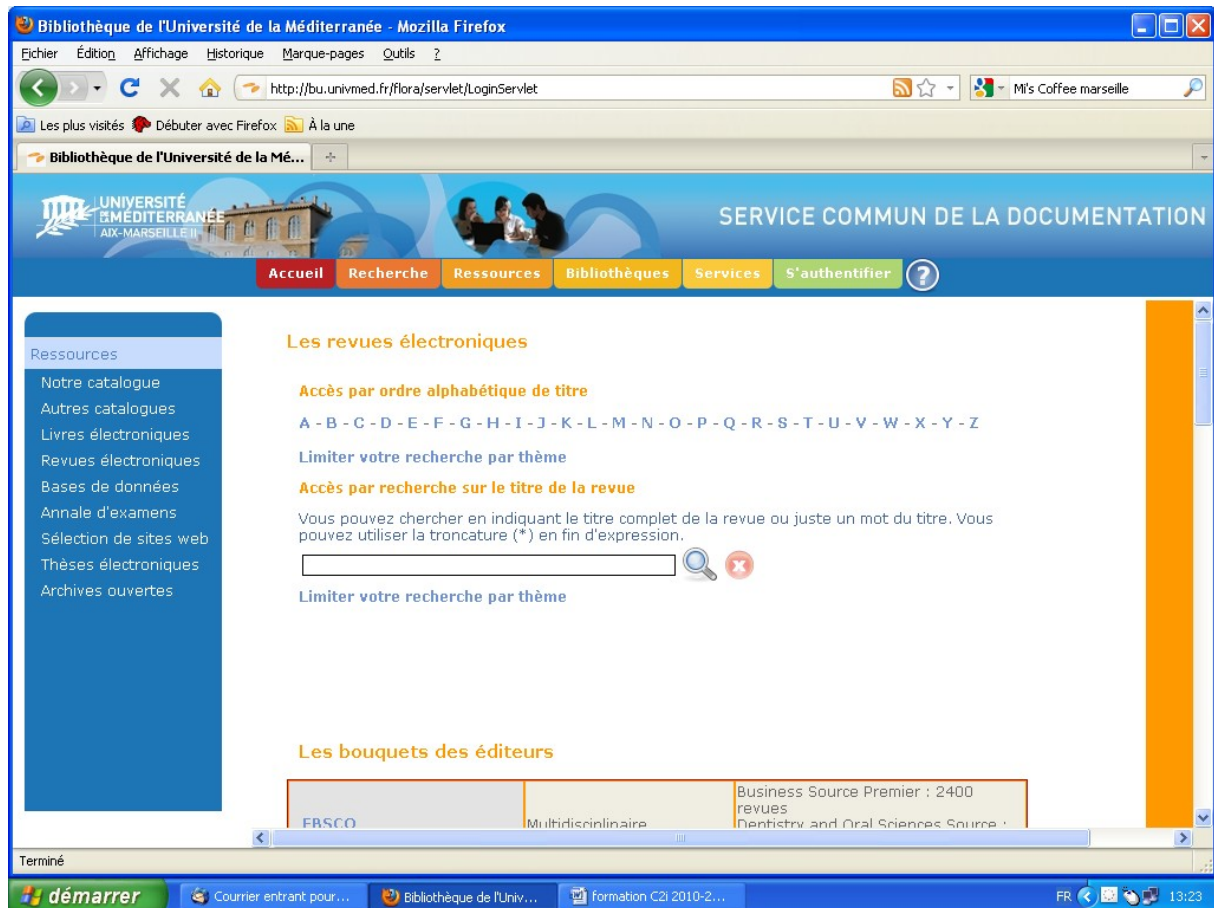
Exercice : Recherchez dans Wikipédia l'article Alan Turing. Cet article vous paraît-il fiable ?

b. Revues électroniques

The screenshot shows the website of the University of Aix-Marseille II's library, titled "Bibliothèque de l'Université de la Méditerranée - Mozilla Firefox". The browser address bar shows the URL <http://bu.univmed.fr/flora/servlet/LoginServlet>. The website header includes the university logo and the text "SERVICE COMMUN DE LA DOCUMENTATION". A navigation menu contains the following items: Accueil, Recherche, Ressources (highlighted), Bibliothèques, Services, and S'authentifier. A left sidebar lists various resources, with "Revue électronique" selected. The main content area displays a grid of resource categories, each with a small image and a square icon:

- Catalogue de la Bibliothèque
- Livres électroniques
- Revue électronique
- Bases de données scientifiques
- Annales d'examens
- Thèses électroniques de l'université
- Archives ouvertes
- Sélection de sites web

The Windows taskbar at the bottom shows the "démarrer" button, several open applications, and the system tray with the date "FR" and time "13:10".



Cette interface de recherche ne permet que de **chercher une revue par son titre** (pas de recherche sujet possible)

On peut accéder au document lui-même (alors que dans un catalogue de bibliothèque on accède à une référence de document mais pas au document lui-même)

Regardez la revue **science** (plus bas sur la page, sous « autres abonnements par thème »)

Possibilité de faire une recherche par sujet (swine flu par exemple)

Possibilité d'ajouter des filtres à la recherche (sous « modify search ») dans le menu de gauche

Tous les articles sont disponibles au format PDF

Regardez le dernier numéro paru (« current issue »)

c. Bases en ligne Web of knowledge

Cette base interroge des périodiques (dépouillés) et on peut donc effectuer une recherche **sujet**

NB : la recherche « topic » interroge les mots du titre, de l'abstract, les mots-clés

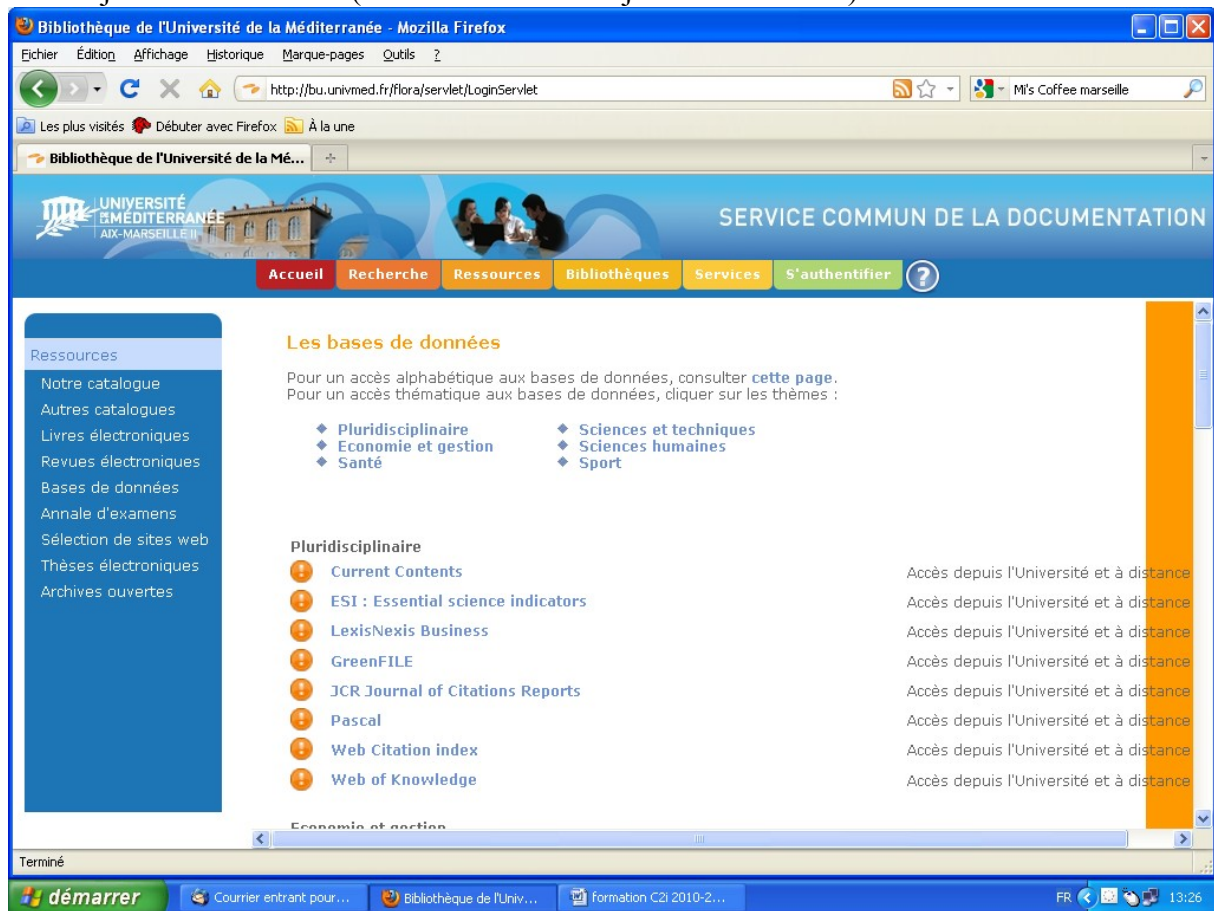
Casse (maj/min) sans importance

Opérateurs booléens : AND OR NOT SAME

Contenu de presque 11 000 périodiques indexés – périodiques choisis sur des critères de qualité (comité de lecture)

Format des périodiques sans importance (papier, électronique, payant, open access etc)

Mise à jour hebdomadaire (souvent effectuée le jeudi ou vendredi)



The screenshot shows the website of the University of Aix-Marseille II's library. The page is titled 'Bibliothèque de l'Université de la Méditerranée - Mozilla Firefox'. The main navigation bar includes 'Accueil', 'Recherche', 'Ressources', 'Bibliothèques', 'Services', and 'S'authentifier'. The 'Ressources' sidebar lists various services like 'Notre catalogue', 'Autres catalogues', 'Livres électroniques', etc. The main content area is titled 'Les bases de données' and provides instructions for accessing the databases. It lists two main categories: 'Pluridisciplinaire' and 'Sciences et techniques'. Under 'Pluridisciplinaire', several databases are listed with their respective access methods:

Database	Access Method
Current Contents	Accès depuis l'Université et à distance
ESI : Essential science indicators	Accès depuis l'Université et à distance
LexisNexis Business	Accès depuis l'Université et à distance
GreenFILE	Accès depuis l'Université et à distance
JCR Journal of Citations Reports	Accès depuis l'Université et à distance
Pascal	Accès depuis l'Université et à distance
Web Citation index	Accès depuis l'Université et à distance
Web of Knowledge	Accès depuis l'Université et à distance

Faites une recherche dans le web of knowledge (swine flu par exemple)

Il s'agit d'une base de données **bibliographiques** (notices non liées au plein texte) – une très grosse base américaine où la moindre recherche ramène un grand nombre de résultats.

NB : certains outils donnent l'accès à des références bibliographiques (catalogue de bibliothèque, bases de données) et d'autres outils donnent l'accès au document lui-même (accès au texte intégral dans certaines bases en ligne)

Regardez la structuration d'une notice (mots-clés etc)

« times cited » : nombre de fois où un article a été cité depuis sa parution vs « reference » : les citations dans l'article

The screenshot displays the ISI Web of Knowledge interface. At the top, the browser title is 'ISI Web of Knowledge [v.4.7] - All Databases Results - Windows Internet Explorer'. The address bar shows a URL starting with 'http://apps.isiknowledge.com/summary.do?tid=1&product=JA&SID=U1blKA9mKdL8jNOK%40C&search_mode=GeneralSearch'. The main content area is titled 'ALL DATABASES Results' with filters for 'Topic=(web 2.0)' and 'Timespan=All Years'. It shows 'Results: 2 072' and 'Page 1 of 208'. A left sidebar titled 'Refine Results' includes sections for 'General Categories' (e.g., SCIENCE & TECHNOLOGY) and 'Subject Areas' (e.g., COMPUTER SCIENCE). The main results list includes:

- Title: Mismatch-based delayed thrombolysis: a meta-analysis. Author(s): Mishra, Nishant K; Albers, Gregory W; Davis, Stephen M, et al. Source: **Stroke** Volume: 41 Issue: 1 Pages: e25-33 Published: 2010 Jan
- Title: Guest Editors' Introduction to Special Issue: Social Networks and Web 2.0. Author(s): Smith-David, J; Wang, B; Westland, JC. Source: **ELECTRONIC COMMERCE RESEARCH AND APPLICATIONS** Volume: 9 Issue: 1 Pages: 3-5 Published: 2010 Times Cited: 0
- Title: Shear Strengthening of RC T-Beams Using Mechanically Anchored Unbonded Dry Carbon Fiber Sheets. Author(s): Galal, K; Mofidi, A. Source: **JOURNAL OF PERFORMANCE OF CONSTRUCTED FACILITIES** Volume: 24 Issue: 1 Pages: 31-39 Published: 2010 Times Cited: 0
- Title: A Dynamic Bridge for Data Sharing on e-Science Grid Implementing Web 2.0 Service. Author(s): Jung, IY; Yeom, HY. Conference Information: **Korea e-Science All Hands Meeting 2008**, Date: SEP 08-09, 2008 Korea Inst Sci & Technol Informat Daejeon SOUTH KOREA. Source: **FUTURE APPLICATION AND MIDDLEWARE TECHNOLOGY ON E-SCIENCE** Pages: 73-83 Published:

Le menu déroulant de gauche permet d'affiner les résultats (en ajoutant des termes, en sélectionnant le type de documents, les auteurs etc)

Les résultats sont en général ordonnés en fonction de la date mais on peut organiser les résultats autrement.

Possibilité de sélectionner les résultats, les envoyer par mail, les traiter avec un logiciel de gestion bibliographique.

Pour trouver éventuellement le document à la bibliothèque, il faut noter le titre de la revue dont il est tiré et entrer ce titre dans la barre de recherche des titres de revues électroniques (sur le site de la BU onglet ressources > revues électroniques)

Ces références sont vérifiées (revues à comité de lecture).