

## Notes pour l'utilisation de la plate-forme CALICO d'analyse de forums

Emmanuel Giguet – François-Marie Blondel  
version du 22 février 2010

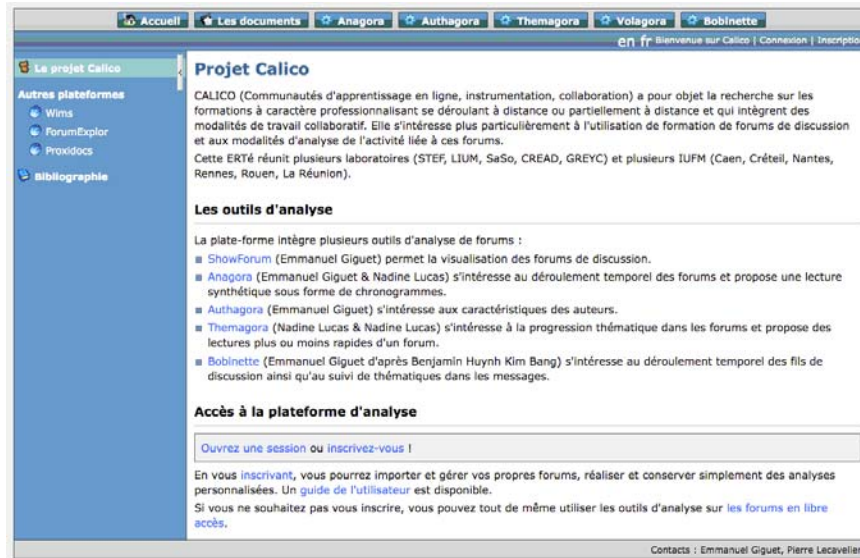
### Sommaire

Présentation.....	2
Espace public et utilisateurs inscrits .....	2
Mode non connecté.....	2
Mode connecté.....	3
Les outils.....	3
Élargir la vue.....	3
Consulter un forum : Showforum .....	3
Restrictions et anonymisation.....	5
Colorier un forum : Colagora .....	5
Filtrer à l'aide des thèmes.....	7
Nommer un thème .....	7
Sauvegarder .....	7
Rafranchir .....	8
Réinitialiser.....	8
Des vues chronologiques et synthétiques : Anagora, Authagora, Themagora .....	8
Une vue sur les fils : Anagora.....	8
Une vue sur les auteurs : Authagora .....	9
Le forum comme un discours : Themagora.....	9
Le forum comme un hypertexte colorié : Bobinette.....	10
Autres actions sur un forum.....	11
Exporter le forum.....	11
Ajouter le forum à la liste des favoris.....	11
Utilisateurs identifiés .....	12
Inscription .....	12
Mon compte .....	12
Gérer les forums .....	12
Forums et thématiques .....	13
Ajouter un forum .....	13
Ajouter ou modifier une thématique.....	14
Historique .....	15
Préparer ses données au format XMLForum.....	15
Avant d'importer un fichier XML au format XMLForum :.....	15
Annexe : Schéma XMLForum.....	17

## Présentation

Cette plate-forme, ainsi que les outils qu'elle rassemble, a été développée dans le cadre de l'ERTé CALICO (Communautés d'apprentissage en ligne, instrumentation, collaboration). Sa mise en place est assurée par Emmanuel Giguet assisté de Pierre Lecavelier et bénéficie des conseils et commentaires des membres de CALICO.

Adresse : <http://crashdump.net/calico/>



La plate-forme donne accès à six outils :

- ShowForum (Emmanuel Giguet) permet la visualisation des forums de discussion.
- Colagora permet de repérer des thématiques et de colorier le contenu d'un forum en accord avec ces thématiques.
- Anagora (Nadine Lucas & Emmanuel Giguet) s'intéresse au déroulement temporel des forums et propose une lecture synthétique sous forme de chronogrammes.
- Authagora (Emmanuel Giguet) s'intéresse aux caractéristiques des auteurs.
- Volagora (Emmanuel Giguet) offre des visualisations quantitatives paramétrables.
- Themagora (Emmanuel Giguet & Nadine Lucas) s'intéresse à la progression thématique dans les forums et propose des lectures plus ou moins rapides d'un forum.
- Bobinette (Emmanuel Giguet d'après Benjamin Huynh Kim Bang) s'intéresse au déroulement temporel des fils de discussion ainsi qu'au suivi de thématiques dans les messages.

Les outils de la plate-forme connaissent deux types de documents, les forums de discussion et les thématiques.

La plate-forme propose aussi une bibliographie sur les outils et le projet CALICO.

## Espace public et utilisateurs inscrits

Selon qu'un utilisateur est inscrit ou non sur la plateforme, les accès aux données seront différents

### Mode non connecté

L'utilisateur non connecté accède uniquement aux forums et aux thématiques déclarées comme `Public` (voir plus loin). Plusieurs exemples de forums sont proposés en accès public, notamment les "OS xxxxx" extraits de la plateforme phpBB du Department of Mathematical and Computer Sciences at the Colorado School of Mines par E Giguet.

Un utilisateur non connecté peut appliquer tous les outils sur ces forums.

## Mode connecté

Une fois identifié, l'utilisateur peut accéder à ses propres forums et thématiques ainsi qu'à ceux qui sont partagés par les autres utilisateurs et qui ont été déclarés comme Calico (voir plus loin, *Utilisateurs identifiés*).


## Les outils

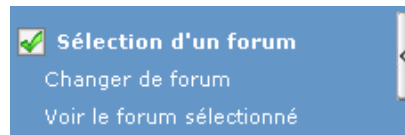
Les outils sont présentés ci-dessous en mode connecté.

On y accède par le bandeau supérieur :




## Élargir la vue

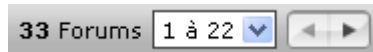
Pour toutes les visualisations, le bouton  situé en haut à droite de la colonne de gauche permet d'élargir l'affichage à toute la largeur de l'écran.



## Consulter un forum : Showforum

Le bouton  affiche la liste des forums et des thématiques.

Le nombre de forums apparaît en bas à droite de la page ; s'il est important, un bouton permet de consulter la suite de la liste sur la page suivante.



La liste est triée par dates de début de forum. On peut trier cette liste sur chaque colonne : Nom du forum, Utilisateur, Accès, Début, Fin, Mess. (nombre de messages) et Aut. (nombre d'auteurs de messages).

	▲ Nom du forum	▲ Utilisateur	▲ Accès	▼ Début	▼ Fin	▲ Mess.	▲ Aut.	
<input type="checkbox"/>	Concertation_avec_les_enseignants	nicole clouet	Groupe	18/09/2008	28/09/2008	17	6	<input type="checkbox"/>
<input type="checkbox"/>	Os Projects 2008	Emmanuel Giguët	Public	01/02/2008	04/05/2008	167	25	<input type="checkbox"/>
<input type="checkbox"/>	tst 00	François-Marie Blondel	Privé	02/12/2007	03/12/2007	3	3	<input type="checkbox"/>
<input type="checkbox"/>	stratégie de communication	nicole clouet	Groupe	22/11/2007	02/03/2008	84	14	<input checked="" type="checkbox"/>
<input type="checkbox"/>	liste E2 2008 Creteil	François-Marie Blondel	Groupe	14/09/2007	04/07/2008	927	37	<input type="checkbox"/>
<input type="checkbox"/>	blog_e1_2008	François-Marie Blondel	Groupe	12/09/2007	18/12/2008	114	38	<input type="checkbox"/>
<input type="checkbox"/>	OS Projects	Emmanuel Giguët	Public	11/09/2007	29/11/2007	168	26	<input type="checkbox"/>
<input type="checkbox"/>	OS Lounge	Emmanuel Giguët	Public	21/08/2007	16/11/2007	171	36	<input type="checkbox"/>
<input type="checkbox"/>	fun	François-Marie Blondel	Privé	12/02/2007	15/02/2007	6	4	<input type="checkbox"/>
<input type="checkbox"/>	besoins et usages2	nicole clouet	Public	06/10/2006	20/11/2006	47	20	<input type="checkbox"/>
<input type="checkbox"/>	absence d'usage = absence de besoin?	Emmanuel Giguët	Public	06/10/2006	11/11/2006	14	11	<input type="checkbox"/>
<input type="checkbox"/>	pagestec 2006 10_12	François-Marie Blondel	Privé	04/10/2006	21/12/2006	77	51	<input type="checkbox"/>

Le nom affiché dans la colonne Nom du forum est un lien actif ; cliquer sur ce nom pour afficher le contenu du forum.

The screenshot shows a forum page titled "OS Concepts". At the top, there are navigation tabs: Accueil, Les documents, Anagora, Authagora, Themagora, Volagora, and Bobinette. On the right, there are links for "Bienvenue sur Calico", "Connexion", "Inscription", "Visualiser", "Exporter", and "Versions".

On the left sidebar, there is a "Sélection d'un forum" section with options to "Changer de forum", "Voir le forum sélectionné" (with "OS Concepts" selected), and "Changer de version". Below it is a "Lancer Anagora" button.

The main content area shows the forum title "OS Concepts" and its creation date "05/12/2007". It states it was created by Emmanuel Giguet and is a forum for the Department of Mathematical and Computer Sciences at the Colorado School of Mines. It contains 244 messages from 22 authors between 30/07/2006 and 28/11/2007, with 77 discussion threads.

There is a filter box for displaying messages "sans restriction sur les auteurs sans anonymisation, sans restriction sur les fils de discussion" between 30/07/2006 and 28/11/2007, with "annuler" and "appliquer" buttons.

Below the filter, there is a navigation bar for "Activer Colagora" with a link to "Afficher les fils de discussion" and a pagination control "1/5 2 3 4 5".

The message list shows the first four messages, all from Mike Colagrosso on 30 Jul 2006:

- The OS announcements and notification topic**: A message about using the forum for announcements and providing a link to an RSS feed.
- Don't email Mike, post here.**: A message about where to post questions to get prompt answers.
- Policy on code posting**: A message about the rules for posting code snippets.
- Post here if you need an Alameda account**: A message about the format for posting user information.

L'affichage par défaut montre le contenu du premier message et la liste des messages suivants par ordre chronologique. Si la liste est trop importante, elle s'affiche sur plusieurs pages que l'on peut faire défiler par les boutons situés avec les options d'affichage :

[Afficher les fils de discussion](#) | [Déplier/replier](#) | <<< 1/5 2 3 4 5 >>>

L'affichage par fils montre la structure des fils de discussion par une indentation.

L'option Tout déplier/replier permet d'afficher le contenu ou seulement le sujet des messages.

This screenshot shows the same forum page but with the "Déplier/replier" option selected, resulting in a threaded view where the content of each message is fully visible.

The navigation bar at the top is identical to the previous screenshot.

The message list is expanded to show the full content of the first four messages:

- The OS announcements and notification topic**: The full text of the message about announcements and the RSS feed link is visible.
- Don't email Mike, post here.**: The full text explaining the purpose of this topic is visible.
- Policy on code posting**: The full text explaining the code posting policy is visible.
- Post here if you need an Alameda account**: The full text explaining the posting format is visible.

The fifth message in the list is dated 21 Aug 2007 and is titled "Post the results of you in-class exercise here".

## Restrictions et anonymisation

### Première méthode :

Cliquer sur les éléments correspondant à la restriction souhaitée, par auteurs, par fils de discussion, ou par dates ou sur l'anonymisation. Les options qui apparaissent permettent de choisir plus finement.

Ce forum n'a aucune restriction sur les auteurs sans anonymisation et n'a aucune restriction sur les fils de discussion entre le 12/01/2006 et le 09/02/2006



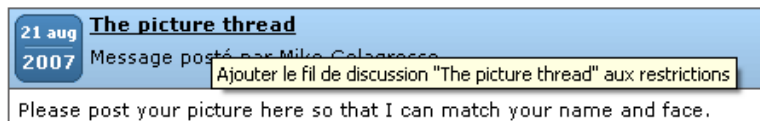
Ainsi pour restreindre à certains auteurs, il suffit de les sélectionner dans la liste déroulante. Cliquer sur Appliquer pour rendre les restrictions effectives. Cliquer sur Annuler supprime toutes les restrictions.

### Seconde méthode :

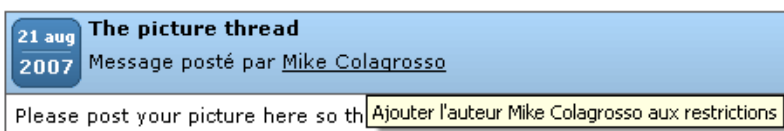
Sur l'affichage d'un message, des bulles d'aides indiquent qu'il est possible de restreindre l'affichage à partir de l'élément pointé :



cliquer sur la date pour restreindre à la date du message affiché,



cliquer sur le sujet du message pour restreindre au fil de discussion,



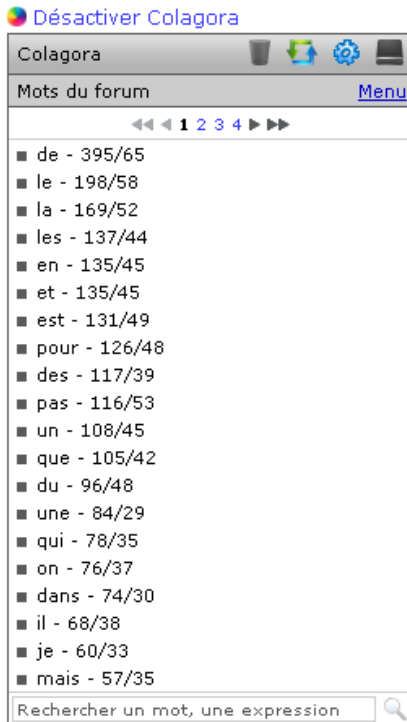
cliquer sur le nom de l'auteur pour restreindre à l'auteur du message.

## Colorier un forum : Colagora

Cet outil est destiné à faciliter la coloration thématique d'un forum. Pour cela, il calcule et affiche les occurrences des mots du forum. Il permet de colorier le forum en sélectionnant des mots particuliers que l'on peut ensuite sauvegarder dans des thématiques.



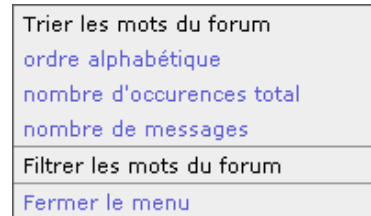
Une fois activé, Colagora affiche la liste des mots les plus fréquents. Le premier nombre est celui des occurrences du mot dans tout le forum, le second indique le nombre de messages dans lequel le mot apparaît.



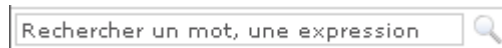
On peut parcourir cette liste par page par page avec les boutons avant et arrière ou les numéros de page.



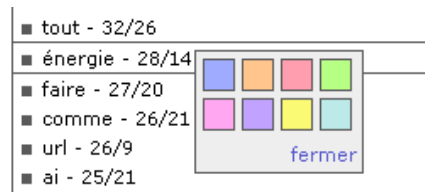
On peut trier les mots de plusieurs manières en activant le [Menu](#) de Colabora.



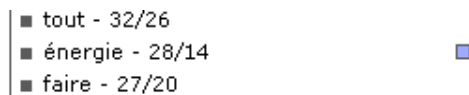
On peut rechercher un mot en indiquant sa racine dans la boîte en bas de la liste :




L'association d'une couleur à un mot se fait en cliquant sur le mot.



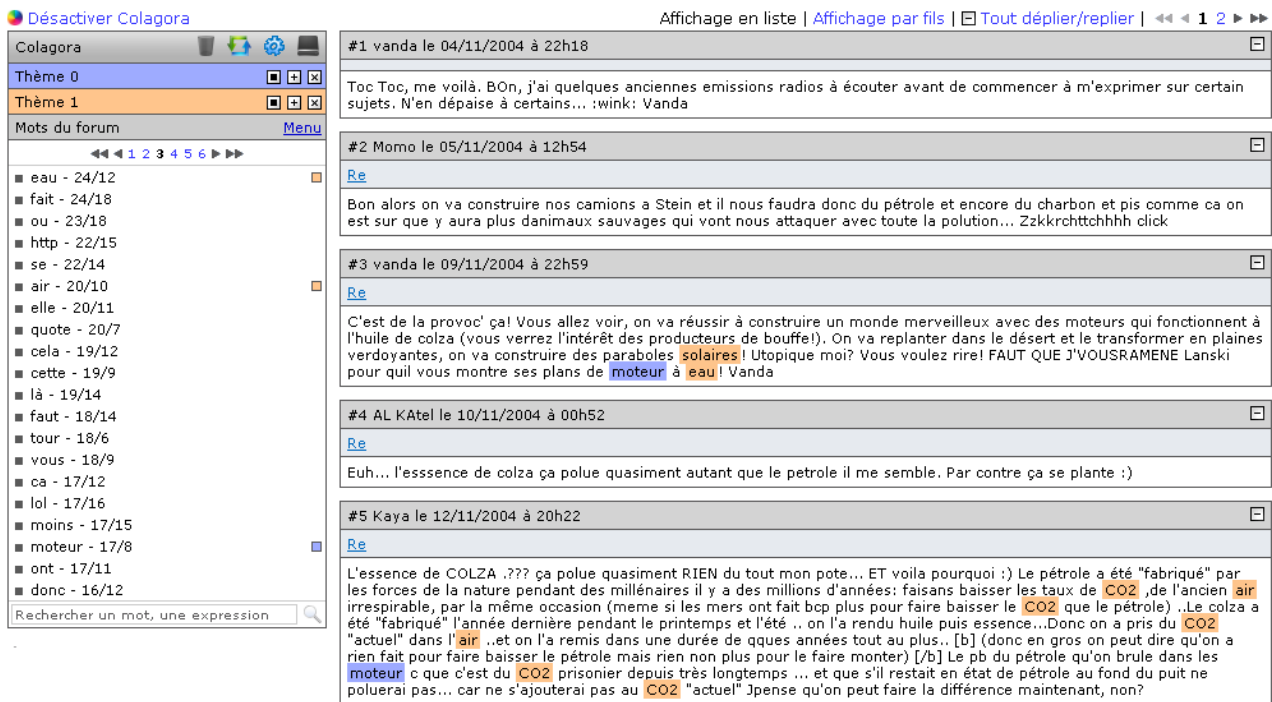
Cliquer sur une couleur pour l'associer au mot.



Dans le bandeau de Colabora, cliquer sur le bouton  pour colorier le forum avec les couleurs associées aux mots.



- Quand plusieurs couleurs sont utilisées, elles apparaissent sous le bandeau de Colagora sous la forme Thème n.

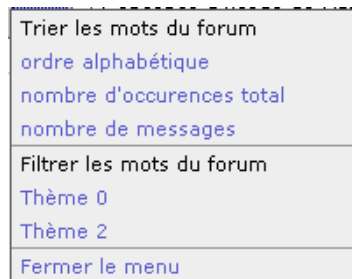


Les boutons qui apparaissent dans le bandeau de chaque thème    permettent de :

- désactiver le thème : il ne provoque plus de coloration
- déplier le thème : les mots associés sont affichés
- supprimer le thème

### Filtrer à l'aide des thèmes

Une des couleurs peut être employée pour éliminer certains mots de l'affichage de la liste de Colagora, comme les mots grammaticaux qui sont très fréquents.




Pour Filtrer les mots du forum, sélectionner le thème contenant les mots à filtrer, leur affichage disparaît de la liste de Colagora.

### Nommer un thème

Cliquer directement sur Thème n pour éditer son nom.



### Sauvegarder

Les associations entre mots et couleurs peuvent être sauvegardées dans une thématique en cliquant sur le bouton . Une thématique rassemble un ou plusieurs thèmes, c'est-à-dire un ou plusieurs groupes de mots qui apparaissent sous la même couleur.

## Rafraichir

Actualise l'affichage des thèmes en tenant compte de tous les mots associés à des couleurs.

## Réinitialiser

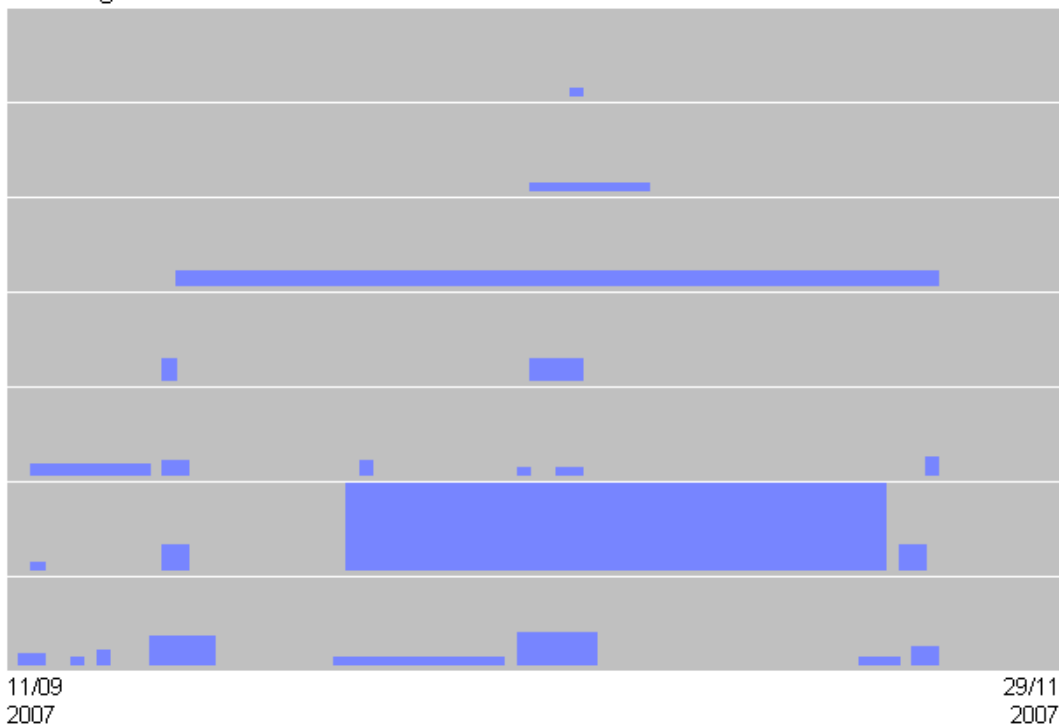
Supprime toutes les associations entre mots et couleurs, excepté celles qui proviennent d'une thématique sélectionnée auparavant dans la liste des thématiques (voir plus loin, *Gérer les forums et les thématiques*).

## Des vues chronologiques et synthétiques : Anagora, Authagora, Themagora

### Une vue sur les fils : Anagora

#### Analyse avec Anagora

7 chronogrammes



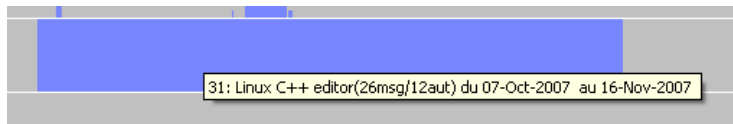
Dans le chronogramme situé au centre de la visualisation, chaque fil est représenté par un bloc rectangulaire coloré (en bleu sur la figure) qui commence avec le premier message du fil et se termine avec le dernier message.

L'axe horizontal représente le temps. Seuls les fils qui contiennent plus de trois messages sont représentés. La longueur du bloc représente la durée totale du fil, et la hauteur du bloc représente le nombre total de messages du fil.

La tranche inférieure commence avec le premier fil. Anagora ajoute une tranche à chaque fois qu'un nouveau fil apparaît en plus des fils déjà actifs. Le nombre de tranches donne ainsi le nombre maximum de fils qui ont été simultanément actifs (7 fils sur la figure).



En passant la souris sur un bloc représentant un fil, on affiche les informations du fil : sujet, nombre de messages, nombre d'auteurs, dates de début et de fin.



## Une vue sur les auteurs : Authagora

Analyse avec Authagora

26 auteurs	13 initiateurs	5 auteurs sans réponse	23 auteurs qui répondent
Mike Colagrosso (31)	IsmaelGonzales (8)	Michael Feineman (1)	Mike Colagrosso (24)
Michael Feineman (20)	Teresa Davies (7)	Mike Colagrosso (1)	smcfarla (13)
IsmaelGonzales (17)	Michael Feineman (5)	rcarpent (1)	Michael Feineman (13)
smcfarla (15)	Mike Colagrosso (5)	Jeremy Norman (1)	AndyMan1 (9)
Teresa Davies (13)	Alex Le* (3)	vsingh (1)	squid314 (6)
AndyMan1 (10)	vsingh (3)		vsingh (5)
vsingh (10)	SeanB* (2)		Corporate Stiff (5)
squid314 (6)	rcarpent* (2)		Kari Macklin (4)
Corporate Stiff (5)	cbrady* (1)		Stuart Fehr* (3)
Brandon Ardiente (4)	AndyMan1* (1)		Adam McCormick* (3)

[Afficher tous les auteurs](#)

master: -

leader: Mike Colagrosso, Michael Feineman

speaker: -

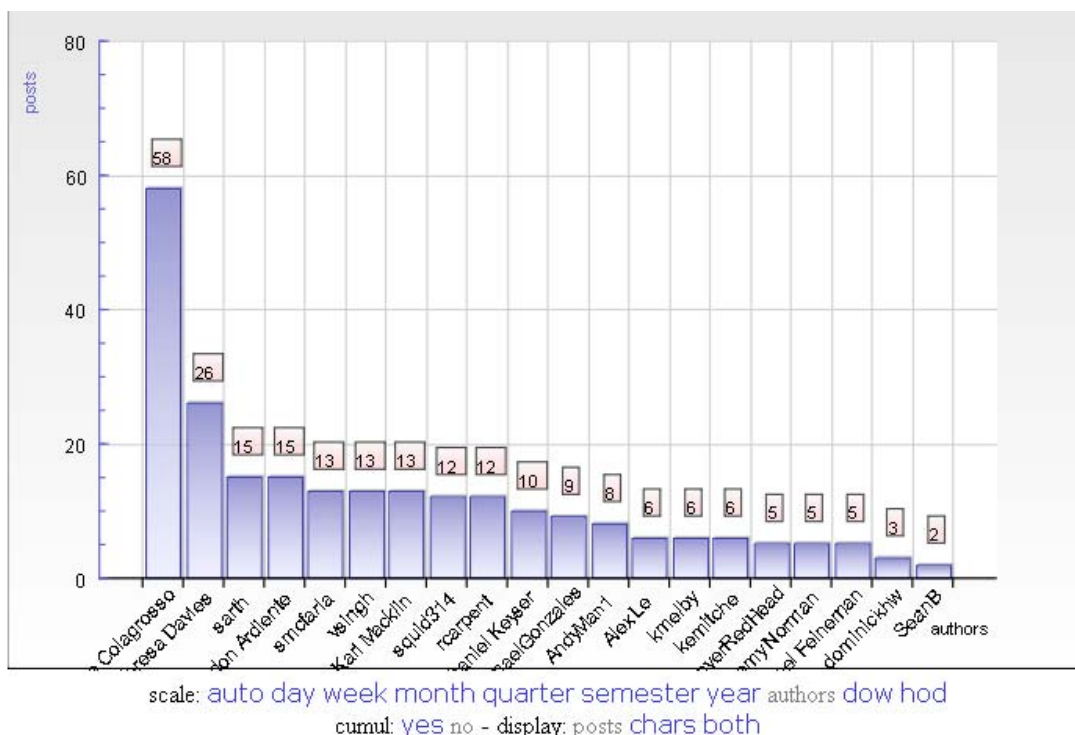
writer: smcfarla, AndyMan1

Le tableau affiche par Authagora comprend en quatre colonnes :

- les auteurs les plus productifs : ceux qui écrivent le plus de messages
- les premiers auteurs : ceux qui initient le plus de fils
- les auteurs sans réponses : ceux qui ont reçu le moins de réponses à leurs messages
- les auteurs qui répondent : ceux qui ont le plus répondu à un message

## Volagora

Cet outil est un outil générique de visualisation des volumes, par dates, périodes, auteurs....



## Le forum comme un discours : Themagora

Themagora considère les messages successifs du forum comme un discours sur lequel on va appliquer des règles d'analyse automatique pour y mettre en évidence des thèmes et des rhèmes.

Chaque nouvelle unité thématique apparaît sur fond bleu ; les boutons et permettent de déplier ou replier les unités. Les et permettent de développer le texte ou de le réduire.

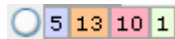
## Le forum comme un hypertexte coloré : Bobinette

Bobinette affiche tous les fils en leur affectant une coloration thématique.

Bobinette ne peut pas fonctionner sans au moins une thématique associée. Vérifiez qu'une thématique a été sélectionnée ou qu'une coloration thématique vient d'être effectuée avec Colagora.

L'affichage des messages du forum s'effectue suivant un axe chronologique horizontal. Les messages appartenant au même fil de discussion sont regroupés sur la même horizontale. S'ils ont été émis le même jour, ils sont aussi regroupés verticalement.

Chaque message est représenté par un cercle suivi du nombre d'occurrences des graphies de chaque thème, avec une coloration d'autant plus saturée que l'occurrence est élevée.

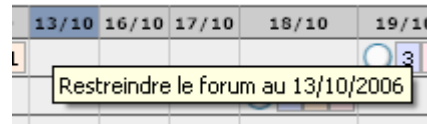


Cliquer sur le cercle pour accéder au contenu du message correspondant. Les mots du message sont coloriés thématiquement. Déplacer le message en le sélectionnant par son bandeau.

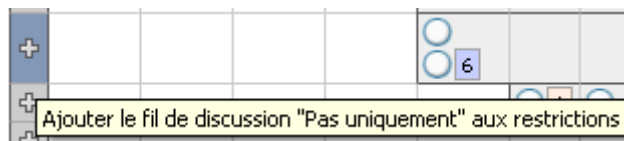
La reconnaissance des graphies tient compte des séparateurs courants. Elle est indépendante de la casse (majuscules/minuscules).

On peut en pratique sélectionner plusieurs thématiques simultanément ; la coloration s'adapte en conséquence.

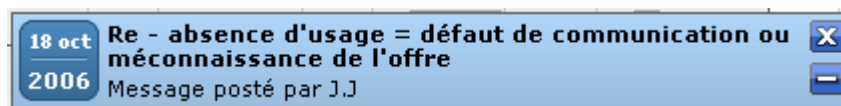
Cliquer sur une date pour restreindre le prochain affichage à cette date.



Cliquer sur un fil - signe en bordure - pour restreindre le prochain affichage à ce fil.



Cliquer sur les éléments du bandeau, date, auteur ou fil pour restreindre le prochain affichage à une date, un auteur ou un fil.

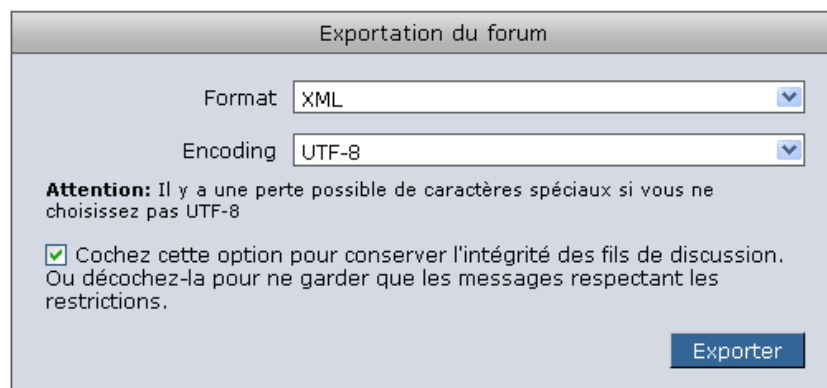


## Autres actions sur un forum

Plusieurs actions peuvent modifier le forum affiché ou son statut.



## Exporter le forum



L'export peut se faire sous deux formats, XMLForum et CSV, et trois encodages de caractères : UTF-8, Latin 1 (ISO 8859-1) et Windows 1251.

Dans le format CSV, le forum est présenté dans une table avec les champs suivants : msgid, datetime, author, subject, msgref, content.

Le format CSV est ensuite réutilisable avec un tableur et particulièrement Open Office Calc qui permet de choisir l'encodage des caractères à l'ouverture du fichier.

## Ajouter le forum à la liste des favoris

Pour retrouver plus facilement les forums sur lesquels on travaille fréquemment.

## Utilisateurs identifiés

### Inscription

Remplir le formulaire en ligne (identifiant et mot de passe).

La validation de l'accès est faite par les administrateurs. Tous les participants au projet CALICO sont acceptés. D'autres participants peuvent être acceptés en fonction de leur intérêt et de leurs travaux en rapport avec de domaine de recherche.

### Mon compte

**Mon compte**

Il permet de voir son profil, ses forums et ses thématiques.

### Gérer les forums et les thématiques

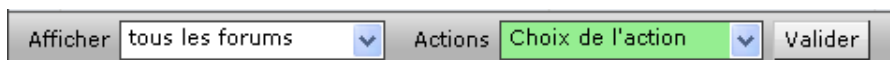
**Les documents**

Il permet de voir la liste des forums et des thématiques.

Utiliser la colonne de gauche pour sélectionner le forum / la thématique que l'on souhaite modifier ou éditer.

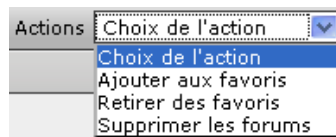
<input type="checkbox"/>	Nom du forum ▲
<input type="checkbox"/>	OS Projects
<input type="checkbox"/>	OS Lounge
<input checked="" type="checkbox"/>	fun
<input type="checkbox"/>	absence d'usage = absence de besoin?
<input type="checkbox"/>	pagestec 2006 10_12
<input type="checkbox"/>	OS Concepts
<input checked="" type="checkbox"/>	pagestec 2006 06_09
<input checked="" type="checkbox"/>	pagestec 2006 03_05

Dans le bandeau inférieur :



La liste **Afficher** permet de choisir entre l'affichage de tous les forums, mes forums ou mes forums favoris ; idem pour les thématiques.

La liste **Actions** permet de modifier, de supprimer ou d'ajouter aux favoris, les forums sélectionnés en colonne de gauche.



**Modifier** permet de changer le nom la description et l'accès (Privé, Calico, Public), et pour les thématiques de modifier directement les thèmes. Vous ne pouvez modifier que les forums que vous avez importé sur la plate-forme.

Modification du forum

Vous pouvez modifier les propriétés du forum :

Nom

Description

Accès

La colonne de droite permet de sélectionner le forum / la thématique sur lequel on va appliquer les outils.

Mess.	▲ Aut.	▲ Sélection
168	26	<input type="checkbox"/>
171	36	<input checked="" type="checkbox"/>
6	4	<input type="checkbox"/>
14	11	<input type="checkbox"/>
77	51	<input type="checkbox"/>

### Forums et thématiques

Le menu situé à gauche donne accès aux forums et aux thématiques, ceux qui sont sélectionnés, ceux qui sont privés, et ceux qui sont favoris. Il permet aussi d'ajouter forum ou thématique.



### Ajouter un forum

Le champ **Nom** est obligatoire.

Le champ **Description** est facultatif. Son contenu est affiché en haut de page quand on demande à voir le forum.

Pour le champ **Accès**, choisir entre les trois valeurs :

Public : accès à tous

Calico : accès restreint aux utilisateurs identifiés uniquement

Privé : accès restreint à l'utilisateur identifié uniquement

Quatre formats de forums sont reconnus, XMLForum, BSCW, MULCE, ASPRM

Ajouter un forum

■ Choisissez d'abord les propriétés du forum

Nom du forum

[Ajouter une description du forum](#)

Accès

■ Vous devez maintenant choisir l'emplacement du forum

► Soit en important un fichier à partir de votre ordinateur

Fichier

► Soit en téléchargeant un fichier à l'adresse indiquée

Adresse internet

[Indiquez un login et un mot de passe](#)

■ Enfin, vous devez choisir le format du fichier

Format  XMLForum  BSCW  MULCE  
 ASPFRM

Activer l'anonymisation des noms

⚠ Cette option sera disponible après l'ajout du forum, mais si vous l'activez dès maintenant, l'anonymisation sera irréversible.

Une option permet de rendre le forum anonyme dès son ajout sur la plate-forme. Cette transformation est irréversible.

### Cas d'un fichier XML

Si la structure et le codage du fichier sont corrects, le contenu du forum s'affiche. La taille du fichier est limitée (limite fixée à 1Mo).

### Ajouter ou modifier une thématique

Trois options permettent de définir une thématique :

- en utilisant les thèmes définis avec Colagora,
- en important un fichier de thématique au format XML,
- en indiquant directement les thèmes et les mots.

Ajouter une thématique

■ Choisissez d'abord les propriétés de la thématique

Nom

Accès

■ Vous devez maintenant en définir le contenu

► [en utilisant Colagora sur un forum](#)

► Soit en important un fichier thématique au format XML

Fichier XML

► Soit en créant vous-même les thèmes et leurs mots-clés

Thèmes

[Agrandir la zone de texte](#)

Dans ce dernier cas, on indique le contenu de la thématique à raison d'un thème par ligne. Chaque ligne comprend : le nom du thème suivi de deux-points suivi des graphies séparées par des virgules.

Exemple d'une ligne dans une thématique :

horaires: dhg, dgh, dotation, horaire, heure, heures

Les graphies peuvent être des mots (horaire) ou des expressions (dotation horaire). La syntaxe des expressions régulières autorise une écriture plus élaborée des mots ou des expressions.

Exemple : `enseign.*?` désigne tous les mots qui commencent par la même racine `enseign`

## Historique

Il permet d'accéder aux derniers traitements effectués et permet de rejouer les actions correspondantes (liens actifs en colonne de gauche).

## Préparer ses données au format XMLForum

Si les données ne sont pas disponibles dans un des formats acceptés, le premier travail consiste à les mettre dans le format XMLForum. Deux solutions sont possibles :

- **Travailler directement sur le fichier XML à l'aide d'un éditeur de texte.**

Cette technique peut être adoptée pour un forum peu volumineux.

Il est plus agréable de travailler avec un éditeur de texte qui fait une coloration syntaxique des fichiers XML (*Notepad2* logiciel libre pour Windows par exemple).

- **Passer par un format table intermédiaire.**

Cette technique facilite la préparation des données; elle est préférable pour les forums un peu volumineux (quelques centaines de ko).

Elle consiste à placer les données dans les colonnes d'une table qui correspondent aux différents champs : id, date, auteur, sujet ...

Le classeur *calico\_table2xmlforum.xls* a été conçu pour la préparation et la conversion au format XMLForum.

En ligne à [http://woops.crashdump.net/calico/resources/xls/calico\\_table2xmlforum.xls](http://woops.crashdump.net/calico/resources/xls/calico_table2xmlforum.xls)

## Avant d'importer un fichier XML au format XMLForum :

Pour éviter les problèmes lors de l'ajout d'un forum, plusieurs vérifications sont conseillées.

### Vérifier l'encodage

Dans le contenu des champs du fichier XMLForum, il faut remplacer tous les caractères interdits (& < >) par un codage acceptable. Le codage standard est le suivant :

&        `&amp;`

<        `&lt;`

>        `&gt;`

Par ailleurs, le contenu des textes doit être compatible avec l'encodage des caractères déclaré en début de fichier XML.

En pratique, deux encodages sont fréquemment utilisés : ISO-8859-15 (Windows par défaut) et UTF-8.

### Tester la validité du fichier

On peut vérifier rapidement la validité en ouvrant le fichier dans un navigateur (Internet Explorer ou Firefox). Mais il est plus efficace d'utiliser un logiciel plus adapté.

### Contrôler quelques points supplémentaires

Vérifier qu'il n'y a pas de champs vides, notamment pour les éléments `<author>` et `<subject>`.

Vérifier précisément le format des dates :

- le format des jours `aaaa-mm-jj` (4caracs-2caracs-2caracs),

la barre peut aussi être utilisée comme séparateur : `aaaa/mm/jj` est autorisé ;

- le format des heures hh:mm  
les secondes ne sont pas prises en compte ;  
tolérance sur les heures : h:mm

Vérifier qu'il n'y a pas de doublons dans les `id` de message.

Vérifier que les liens sur le message d'origine du fil de discussion (`msgref`) pointent bien sur un message (`id`) déjà présent dans le fichier et qu'il n'y a pas de références circulaires d'un message qui pointe sur lui-même.

Ces deux dernières vérifications peuvent être faites avec le classeur *calico\_table2xmlforum*.



## Annexe : Schéma XMLForum

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
  <xs:element name="forum">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" ref="message"/>
      </xs:sequence>
      <xs:attribute name="name" use="required" />
    </xs:complexType>
  </xs:element>
  <xs:element name="message">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="header"/>
        <xs:element ref="body"/>
      </xs:sequence>
      <xs:attribute name="id" use="required" type="xs:integer"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="header">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="datetime"/>
        <xs:element ref="author"/>
        <xs:element ref="subject"/>
        <xs:element minOccurs="0" ref="msgref"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="datetime" type="xs:string"/>
  <xs:element name="author" type="xs:string"/>
  <xs:element name="subject" type="xs:string"/>
  <xs:element name="msgref">
    <xs:complexType>
      <xs:attribute name="id" use="required" type="xs:integer"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="body">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="content"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="content" type="xs:string"/>
</xs:schema>
```