
SFer

Guide de l'utilisateur

Documentation basée sur SFer 2.0beta

28 juin 2007

Contact : Fremen <fremen@planet-shitfliez.org>

Table des matières

1	Présentation générale	4
1.1	Avantages et inconvénients	4
1.1.1	Avantages de SFer	4
1.1.2	Inconvénients de SFer	5
1.2	Structure globale	5
1.3	Description des différents composants	5
1.3.1	web2sql	7
1.3.2	SFerConfig	7
1.3.3	SFer	7
2	Installation et configuration	8
2.1	Configuration requise	8
2.2	web2sql	8
2.2.1	Prérequis	8
2.2.2	Compilation et installation	9
2.2.3	Configuration	9
2.2.3.1	Copie du fichier de configuration	9
2.2.3.2	Structure du fichier de configuration	9
2.2.3.3	Orthographe des mois (month_config)	10
2.2.3.4	Informations spécifiques à un forum (forum_specific_info)	10
2.2.3.5	Expressions régulières (parser_config)	11
2.3	SFer	12
2.3.1	Prérequis	12
2.3.2	Compilation et installation	13
2.3.3	Configuration	13
2.3.3.1	Copie du fichier de configuration	13
2.3.3.2	Configuration de SFer : sfer_config.xml	13
2.3.3.3	Configuration des plugins	14
2.3.3.4	Création de la base de données	15
2.4	SFerConfig	16
2.4.1	Prérequis	16
2.4.2	Installation et configuration	16
2.4.2.1	Paramétrage	16

TABLE DES MATIÈRES

3	Utilisation	17
3.1	web2sql	17
3.1.1	Exécution	17
3.1.2	Informations en sortie et fonctionnement	17
3.2	SFer	18
3.2.1	Exécution	18
3.2.2	Paramètres	19
3.2.3	Informations en sortie et fonctionnement	20
	3.2.3.1 Analyse des messages	20
	3.2.3.2 Génération des statistiques	20
3.3	SFerConfig	21
3.3.1	Partie utilisateur	21
3.3.2	Partie modération	23
3.3.3	Partie administration	23
4	A propos	24
4.1	Licence	24
4.2	Contacteur l'auteur	24

Introduction

Développé en C et disposant d'une interface en PHP, SFer est un programme permettant de générer des statistiques détaillées de forums de discussion.

Ce document détaille toutes les caractéristiques de ce logiciel dans la partie [1](#). Cette partie n'est pas seulement informative. En effet, l'installation et la configuration de SFer peut être assez lourde, par conséquent il est préférable d'avoir une idée plus claire de ce programme avant de débiter l'installation.

Ce guide explique également comment installer, configurer et utiliser SFer, ainsi que les outils disponibles en complément. Cette procédure nécessite de faire de nombreuses manipulations et il est donc fortement recommandé de lire en détail les parties correspondantes lors d'une première installation et utilisation.

Enfin, toute information complémentaire est disponible dans la partie [4](#).

1 Présentation générale

Il existe déjà divers scripts et logiciels capables de générer des statistiques de forums. Certains se présentent sous la forme d'une extension (par exemple pour des forums phpBB), d'autres sont intégrés par défaut au système chargé de gérer les messages d'utilisateurs.

SFer se distingue de ces autres logiciels par le fait qu'il peut s'adapter à différents types de forums. Il peut se connecter à la base de données MySQL contenant les messages, quelle que soit sa structure, tout comme il est possible de le faire parcourir les différentes pages d'un forum pour lui faire récupérer les contributions de chaque visiteur.

Avant de décrire plus en détail l'architecture et le fonctionnement de ce logiciel, une première section présente la liste des avantages et inconvénients propres à SFer, pour vous aider à savoir s'il répond vraiment à vos besoins.

1.1 Avantages et inconvénients

1.1.1 Avantages de SFer

- **Flexibilité**
 - Capable de se connecter à n'importe quelle base de données MySQL contenant la liste des messages
 - Possibilité de parcourir automatiquement toutes les pages d'un forum (par http) à l'aide d'un outil fourni (web2sql). Cela rend possible la création de statistiques pour des forums dont l'accès à la base de données n'est pas autorisé.
 - La lecture des données comme l'écriture des statistiques est gérée par des plugins. Il est donc possible d'écrire son propre plugin en cas de besoin bien spécifique.
- **Supporte des forums de taille importante et en grand nombre**
 - La consommation mémoire est constante (SFer s'appuie sur MySQL dont la gestion de mémoire peut être réglée et limitée sans que cela n'influe fortement sur les performances)
 - Un paramètre permet de créer une base de données par forum plutôt que d'utiliser une base de données générale et unique. Cela évite qu'un grand nombre de forums soit synonyme de ralentissements.
 - SFer enregistre les calculs réalisés d'une session à l'autre. Ainsi, lors d'une mise à jour

1.2. STRUCTURE GLOBALE

- des statistiques, seule l'analyse des nouveaux messages est nécessaire.
- **SFer est un logiciel libre**
 - Toute partie du code peut être adaptée en fonction des besoins.

1.1.2 Inconvénients de SFer

- **Lourd à mettre en place et à configurer**
 - Rendement temps d'installation/utilité limité si l'objectif est de faire des statistiques pour un ou deux petits forums
- **Quelques connaissances techniques nécessaires pour configurer SFer**
 - Savoir écrire des expressions régulières (si les pages du forum doivent être téléchargées par http)
 - Savoir écrire des requêtes MySQL (si SFer doit accéder à la base de données contenant les messages du forum)

1.2 Structure globale

De façon générale, SFer comprend trois principaux composants :

- **SFerConfig**, une interface PHP pour réaliser différents réglages tant d'un point de vue utilisateur qu'administrateur
- **web2sql**, un composant facultatif capable de télécharger le contenu d'un forum de discussion disponible sur internet via http et de stocker les données récoltées dans une base de données MySQL
- **SFer**, chargé de lire les nouveaux messages présents dans une base de données source, et les analyser et de générer les statistiques correspondantes

Les interactions entre ces trois composants sont représentées sur la figure 1.1 page 6.

À cela vient d'ajouter deux types de bases de données :

- **la base de données propre à SFer** contenant :
 - les options de configuration gérées par SFerConfig
 - les données obtenues par l'analyse des messages des forums et évitant à SFer de réanalyser tous les messages lors de chaque mise à jour des statistiques
 - les comptes utilisateur
- **une base de données externe** contenant les messages du forum à analyser

1.3 Description des différents composants

Cette section présente plus en détail chaque composant cité précédemment, de sorte à expliquer plus précisément leur fonctionnement.

1.3. DESCRIPTION DES DIFFÉRENTS COMPOSANTS

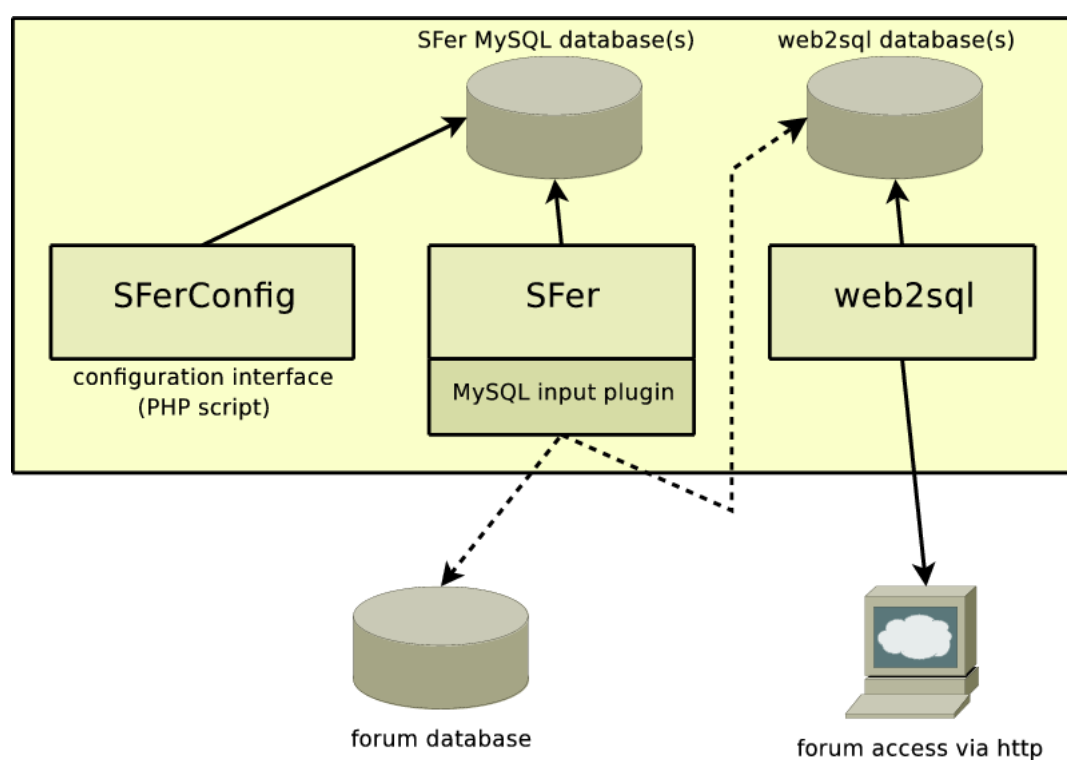


FIG. 1.1 – Architecture globale de SFer

1.3.1 web2sql

web2sql est un logiciel totalement indépendant de SFer. Il est utilisé pour faire des statistiques de forums de discussion dont la base de données n'est pas directement accessible. Pour ce faire, il va télécharger les messages directement sur le site concerné, via http, comme nous le ferions avec un navigateur internet.

En pratique, un fichier de configuration permet de décrire la structure du site :

- le format des liens (quel est le lien vers la page p du topic t ?)
- la format des pages (où se trouve l'auteur du quatrième message sur la page courante ? Où se trouve le n^{eme} message de la page ?)
- les types de balises à interpréter. Par exemple, les smileys se présentent souvent sous forme de balises . Il faut pouvoir les récupérer

Ces points sont évidemment détaillés dans la partie 2, qui traite la configuration de chaque composant.

Une fois que ces informations sur le site à parcourir ont été correctement définies, web2sql va télécharger tous les nouveaux messages du ou des forums définis et les stocker dans sa propre base de données. Celle-ci pourra ensuite être accédée par SFer pour générer les statistiques.

1.3.2 SFerConfig

SFerConfig est une interface PHP qui permet à l'administrateur de configurer SFer (choix des forums à traiter, options des statistiques, etc) et aux utilisateurs de pouvoir s'inscrire et personnaliser leurs informations à prendre en compte dans les statistiques.

Tout cela est présenté plus en détail dans la section 3.3 page 21.

1.3.3 SFer

SFer est un programme en ligne de commande. En entrée, il accède à une base de données contenant les messages du ou des forums de discussion. Les nouveaux messages sont analysés et les informations qui en résultent sont stockées dans la base de données interne à SFer. Ensuite, en sortie, les statistiques des forums sont générées, en fonction des paramètres définis par SFerConfig.

SFer s'appuie sur un système de plugins en entrée comme en sortie. A l'heure actuelle, seul un plugin d'entrée et un plugin de sortie sont disponibles, respectivement pour l'accès à une base de données MySQL et la génération de fichiers HTML. Néanmoins, cela peut faciliter la modification de ce programme pour une utilisation plus spécifique.

2 Installation et configuration

2.1 Configuration requise

La configuration requise ci-dessous a été définie comme une configuration en dessous de laquelle une forte dégradation des performances est à prévoir.

- **Processeur** : 800MHz
- **Mémoire** : 256Mo
- **Espace disque** : prévoir un espace supérieur ou égal à la somme des tailles de chaque message, et deux fois cet espace si les messages doivent être rapatriés avec web2sql

De façon générale, SFer s'appuie fortement sur une base de données MySQL pour stocker les informations sur les messages analysés. Par conséquent, les performances de SFer vont être directement liées à celles de MySQL.

À noter aussi que SFer ne nécessite pas de réanalyser toutes les messages d'un forum à chaque mise à jour de statistiques. Donc même dans les cas où l'analyse serait un peu longue, elle sera faite une bonne fois pour toutes.

2.2 web2sql

Vous pouvez passer cette section si vous comptez régler SFer pour qu'il accède directement à la base de données contenant les messages à analyser.

2.2.1 Prérequis

web2sql s'appuie sur les bibliothèques suivantes :

- libparsifal ¹
- pcre ²
- libmysqlclient ³

¹Site de libparsifal

²Site de PCRE

³paquet présent dans la plupart des distributions Linux, avec un nom similaire à libmysql-client-devel

2.2. WEB2SQL

– libcurl ⁴

Tout cela doit être installé avant de procéder à la compilation.

2.2.2 Compilation et installation

La compilation et l’installation de web2sql se font de façon classique :

```
./configure
make
make install # (en root)
```

2.2.3 Configuration

2.2.3.1 Copie du fichier de configuration

web2sql va chercher son fichier de configuration dans le répertoire `$HOME/.web2sql`

Pour procéder à la configuration de web2sql, copions le fichier exemple fourni par défaut dans le répertoire utilisateur :

```
mkdir ~/.web2sql
cp /usr/local/etc/web2sql_config.xml ~/.web2sql
```

2.2.3.2 Structure du fichier de configuration

Nous allons maintenant éditer ce fichier (`~/.web2sql/web2sql_config.xml`) avec votre éditeur préféré.

Voici sa structure :

- Au début, la DTD (la “syntaxe” du fichier XML). Elle ne doit pas être modifiée.
- Ensuite, l’orthographe des mois tels qu’ils peuvent être écrits sur le site où les messages seront téléchargés
- Un peu plus loin, une liste de blocs `forum_specific_info`, chacun définissant les informations propres à un forum de discussion
- Ensuite, des expressions régulières décrivant où et comment extraire les données de chaque page.
- Enfin, les modifications à apporter sur le texte téléchargé (suppression de balises, etc.)

⁴[Site de cURL et libcurl](#)

2.2. WEB2SQL

2.2.3.3 Orthographe des mois (month_config)

Les mois peuvent être écrits de façon très différente d'un forum à l'autre. Certains les abrègent en trois lettres, d'autres utilisent leur nom anglais, certains utilisent les accents, d'autres non, etc.

La section `month_config` permet de définir toutes les syntaxes souhaitées.

2.2.3.4 Informations spécifiques à un forum (forum_specific_info)

Contenu de chaque bloc Pour chaque bloc `forum_group`, il est possible de définir un ou plusieurs forums, chacun étant représenté par un sous-bloc `forum_specific_info`. Les informations suivantes doivent y être définies :

- Bloc `forum_config`
 - `forum_num` : identifiant du forum sur le site qui l'héberge. Cet identifiant doit se trouver quelque part dans les liens vers n'importe quelle page du forum concerné. Une façon de le repérer est d'aller dans la section du site proposant la liste des forums, et de comparer les liens vers chacun d'entre eux. En général, une valeur change d'un forum à l'autre : c'est l'identifiant de forum.
 - `forum_name` : nom du forum
 - `topic_url` : URL permettant d'accéder à la liste des topics pour le forum concerné. Étant donné que cette liste peut être répartie sur plusieurs pages, une expression arithmétique encadrée par le symbole \$ doit être écrite pour que web2sql puisse parcourir chaque page (voir le paragraphe suivant pour plus de détails)
 - `post_url` : URL permettant d'accéder à une page donnée dans un topic donné. Là aussi, certaines parties de l'URL sont variables : le numéro de page et le numéro de topic. Ces parties variables doivent être décrites par des expressions arithmétiques (voir le paragraphe suivant pour plus de détails)
- Bloc `mysql_config`
 - `mysql_hostname` : machine contenant la base de données MySQL où web2sql stockera les messages téléchargés
 - `mysql_login` : login pour l'accès à MySQL
 - `mysql_password` : mot de passe pour l'accès à MySQL
 - `mysql_dbname` : nom de la base de données où web2sql stockera les données de ce forum. Si elle n'existe pas, web2sql la créera automatiquement
 - `mysql_port` : port du serveur MySQL

Expressions arithmétiques Les URL définies dans le bloc `forum_config` dépendant d'un numéro de page ou encore d'un numéro de topic, des variables doivent être insérées.

Néanmoins, le concept de numéro de page change d'un forum à l'autre. Par exemple, dans les forums phpBB, la position dans un topic va être définie par l'argument `start`, qui correspond au numéro du premier message affiché en haut de page, moins un. Autrement dit, s'il y a

2.2. WEB2SQL

20 messages par page, la page 1 correspondra à `start=0`, la page 2 à `start=20`, etc.

Il faut donc pouvoir faire la conversion entre le numéro de page et la valeur effectivement utilisée par le forum de discussion. C'est là qu'entrent en jeu les expressions arithmétiques.

Une expression arithmétique est composée de nombres entiers naturels (nombres sans virgule supérieurs ou égaux à 0), d'opérateurs ...

- + pour l'addition
- - pour la soustraction
- * pour la multiplication
- / pour la division
- *mod* pour le modulo
- (et) pour encadrer des sous-expressions à calculer en priorité

... et de variables :

- `numpage` : numéro de la page attendue par `web2sql` (commence à 1 et avance jusqu'à ce qu'il n'y ait plus d'informations à lire)
- `numpagez` : raccourci pour `numpage - 1`
- `numtopic` : numéro du topic que `web2sql` souhaite parcourir (valable uniquement pour le paramètre `post_url`)

Pour écrire une expression arithmétique dans une URL, il suffit de l'encadrer par un \$ de chaque côté. Pour reprendre l'exemple du forum phpBB, l'URL aura une forme similaire à :

`http://www.web-site.org/viewtopic.php?t=$numtopic$&start=$numpagez*20$`

2.2.3.5 Expressions régulières (`parser_config`)

Le bloc `parser_config` se divise en trois parties :

- `parser_topic` : indique à `web2sql` comment extraire les informations contenues dans une page listant les topics
- `parser_post` : indique à `web2sql` comment extraire les informations contenues dans une page listant les messages d'un topic donné
- `parser_replace` : permet de remplacer / supprimer des expressions dans le message téléchargé par `web2sql`

`parser_topic` et `parser_post` `parser_topic` et `parser_post` ont exactement la même structure.

Un premier paramètre, `topic_search_order` ou `post_search_order` permet d'indiquer dans quel sens sont affichés les topics ou les messages. Le mode "normal" (l'ordre chronologique en allant du haut vers le bas) est généralement utilisé. L'autre mode a été mis en place pour l'écriture d'une extension éventuelle pour des blogs.

Ensuite suivent un ou plusieurs blocs `search_topic` ou `search_post`. S'ils sont plusieurs, ces blocs seront exécutés dans le même ordre par `web2sql` (du premier au dernier). Chaque bloc permet de dire, à l'aide d'une expression régulière (`parser_*_regex`), où se trouvent les

2.3. SFER

informations recherchées par web2sql (numéro de topic, nom du topic, nombre de réponses, auteur, etc).

Chaque expression régulière peut contenir des sous-expressions, encadrées par des parenthèses (). La première sous-expression est la numéro 1, la deuxième la numéro 2, etc. Les balises qui suivent l'expression régulière permettent de dire à quoi correspond chaque sous-expression. Par exemple, `<parser_topic_author>4</parser_topic_author>` signifie que la sous-expression numéro 4 correspond au nom de l'auteur du topic.

En pratique, web2sql va parcourir chaque bloc `search_*`, rechercher sur la page courante l'expression régulière donnée (et ce un nombre de fois défini par `parser_*_search_count`), et enregistrer les sous-expressions correspondantes, lorsqu'il y en a. Une fois l'ensemble des blocs `search_*` parcourus, web2sql va vérifier qu'il a trouvé autant de dates que de noms d'auteur, autant de noms de topics que d'identifiants, et plus généralement que le nombre d'occurrences pour chaque catégorie de valeurs est égal.

Un exemple est inclus dans le fichier `web2sql_config.xml` et a été testé sur un vieux forum phpBB. A noter que l'ordre des balises ne doit pas être changé et que toutes les balises ne sont pas obligatoires. Par exemple, si l'on a une sous-expression indiquant le nombre de réponses dans un topic donné, la balise correspondant au nombre de messages (et non de réponses) dans un topic du topic n'aura pas lieu d'être utilisée.

parser_replace `parser_replace` peut être ensuite utilisé zéro, une ou plusieurs fois. Chaque bloc définit une expression à rechercher (`parser_replace_regex`), et cette expression sera remplacée par une chaîne de caractères définie par `parser_replace_subst`.

En pratique, cette fonctionnalité peut servir à remplacer des balises `` par des codes de smileys qui seront définis et reconnus dans SFer, à supprimer d'autres balises ou à remanier le contenu des messages.

2.3 SFer

2.3.1 Prérequis

SFer s'appuie sur les bibliothèques suivantes :

- libaspell-devel⁵
- libparsifal⁶
- libmysqlclient⁷

Tout cela doit être installé avant de passer à la suite

⁵correcteur orthographique disponible sur le [site de GNU Aspell](#)

⁶[Site de libparsifal](#)

⁷paquet présent dans la plupart des distributions Linux, avec un nom similaire à `libmysql-client-devel`

2.3.2 Compilation et installation

La compilation et l'installation de SFer se font de façon classique :

```
./configure
make
make install # (en root)
```

2.3.3 Configuration

2.3.3.1 Copie du fichier de configuration

SFer va chercher son fichier de configuration dans le répertoire `$HOME/.sfer`

Pour procéder à la configuration de SFer et de ses plugins, copions les fichiers exemples fournis par défaut dans le répertoire utilisateur :

```
mkdir ~/.sfer
cp /usr/local/etc/sfer_config.xml ~/.sfer
cp /usr/local/etc/input_mysql_config.xml ~/.sfer
cp /usr/local/etc/output_html_out_config.xml ~/.sfer
```

2.3.3.2 Configuration de SFer : `sfer_config.xml`

La majeure partie des paramètres de SFer pourra être définie via l'interface PHP, `SFerConfig`. Néanmoins, certains points doivent être précisés dans le fichier de configuration `sfer_config.xml` :

- Bloc `mysql_config`
 - `mysql_hostname` : machine contenant la base de données MySQL où SFer stockera ses données
 - `mysql_login` : login pour l'accès MySQL
 - `mysql_password` : mot de passe pour l'accès MySQL
 - `mysql_dbname` : nom de la base de données utilisée par sfer
 - `mysql_port` : port du serveur MySQL
- `input_plugin` : plugin utilisé par SFer pour récupérer les nouveaux messages à analyser (actuellement l'unique plugin est `libsfer_mysql.so`)
- `output_plugin` : plugin utilisé par SFer pour générer les statistiques (actuellement l'unique plugin est `libsfer_htmlout.so`)
- Bloc `parser_config`
 - `aspell_language` : langage utilisé par le correcteur orthographique Aspell. `fr_FR` désigne le français.
 - `min_word_length` : longueur minimale d'un mot. En dessous de cette longueur, le mot sera comptabilisé mais ne sera pas ajouté dans le dictionnaire (gain de temps de calcul et permet d'éliminer les mots trop courants)

2.3. SFER

- `topic_words` : une valeur non nulle active la création d'un dictionnaire pour chaque topic. Cela alourdit de façon significative l'analyse des messages.
- `split_db` : une valeur non nulle active la création automatique d'une sous-base de données pour chaque nouveau forum. Cette fonctionnalité permet d'augmenter les performances quand un grand nombre de forums est pris en compte. Une fois que SFer a commencé sa première analyse de messages, cette option ne doit plus être modifiée. Dans le cas contraire, l'application pourrait avoir un comportement aléatoire.

2.3.3.3 Configuration des plugins : `input_mysql_config.xml` et `output_html_config.xml`

`input_mysql_config.xml` Ce fichier configure le plugin d'entrée, chargé de récupérer les messages à analyser depuis une base de données MySQL.

Chaque bloc `forum_info` détaille les requêtes MySQL à réaliser (sous-bloc `select_info`) pour un ensemble de forums donnés (chacun étant détaillé dans un sous-bloc `forum_specific_info`).

Quatre requêtes doivent être définies dans `select_info` :

- `select_post_count` : retourne le nombre total de messages disponibles dans la base de données
- `select_post_group` : retourne un ensemble de messages
 - les colonnes retournées doivent être, dans l'ordre :
 1. l'identifiant du message (NULL s'il n'y en a pas)
 2. l'identifiant du topic auquel le message appartient
 3. la position du message dans le topic (0 = début)
 4. la date et heure (type datetime attendu)
 5. le pseudonyme de l'auteur du message
 6. la valeur 1 si l'auteur a un statut d'administrateur, 0 s'il ne l'est pas, ou NULL si l'information n'est pas disponible
 7. la valeur 1 si l'auteur a un statut de modérateur, 0 s'il ne l'est pas, ou NULL si l'information n'est pas disponible
 8. le corps du message
 9. l'identifiant du pseudonyme de l'auteur (NULL s'il n'y en a pas)
 10. le nom du topic où le message a été posté
- la requête doit être écrite de sorte à ce que l'ordre des messages ne change pas d'une fois à l'autre, et à ce que les nouveaux messages se trouvent toujours après les anciens. En effet, SFer ne relit pas tous les messages à chaque fois et se contente de lire les messages "à la fin" de la requête
- la requête doit se terminer par `LIMIT %d, %d` pour permettre à SFer de sélectionner uniquement les lignes dont il a besoin.
- `select_random_post_forum` : cette requête n'est pas utilisée à l'heure actuelle et peut être ignorée

2.3. SFER

- `select_random_post_topic` : retourne un message dont
 - l'identifiant du topic auquel il appartient se trouve dans la première occurrence de %d
 - la position du message dans le topic se trouve dans la deuxième occurrence de %d (position de début = 0)
 - les colonnes retournées sont les mêmes que dans `select_post_group`, sauf que le nom du topic n'a pas besoin d'être retourné

Remarque : les requêtes fournies par défaut correspondent exactement à la structure des bases de données générées par web2sql. Donc si vous utilisez web2sql, elles n'ont pas besoin d'être modifiées.

Pour chaque bloc `forum_info`, il est possible de définir un ou plusieurs forums, chacun étant représenté par un sous-bloc `forum_specific_info`. Les informations suivantes doivent y être définies :

- Bloc `forum_config`
 - `forum_num` : identifiant du forum
 - `forum_name` : nom du forum
- Bloc `mysql_config`
 - `mysql_hostname` : machine contenant la base de données MySQL où se trouvent les messages à rapatrier
 - `mysql_login` : login pour l'accès à MySQL
 - `mysql_password` : mot de passe pour l'accès à MySQL
 - `mysql_dbname` : nom de la base de données où se trouvent les messages à rapatrier
 - `mysql_port` : port du serveur MySQL

output_html_config.xml Le fichier `output_html_config.xml` permet de choisir le répertoire par défaut où les statistiques au format HTML seront générées, via le paramètre `output_directory`.

Le répertoire fourni est censé être créé et les droits en écriture doivent être donnés à l'utilisateur qui générera les statistiques.

2.3.3.4 Création de la base de données

Une fois les fichiers de configuration de SFer mis en place, il reste encore à créer une base de données vide qui sera remplie par SFer au fur et à mesure. Pour ce faire, il suffit de lancer SFer avec le paramètre `-a` :

```
$ sfer2 -a
SFer 2.0beta
Written by Fremex <fremex@planet-shitfliez.org>
```

```
This is free software; see the source for copying conditions.
There is NO warranty;
not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```


2.4. SFERCONFIG

Main SFer database 'sfer' created

You can now use SFerConfig PHP interface to use the default account (login = admin, pass = admin). Don't forget to change your password.

Si tout s'est bien passé, un message similaire à celui ci-dessus sera indiqué.

2.4 SFerConfig

2.4.1 Prérequis

SFerConfig nécessite l'installation d'un interpréteur PHP et d'un serveur web (par exemple Apache).

2.4.2 Installation et configuration

L'intégralité du répertoire SFerConfig doit être copié sur le serveur web.

2.4.2.1 Paramétrage

Le fichier `common.php` contient tous les principaux paramètres de SFerConfig qui doivent être modifiés. Voici leur description :

- `site_name` : Nom du site auquel cette interface est rattachée
- `support_link` : Cette adresse sera fournie à chaque fois qu'un problème se produira. Cela peut être un forum de discussion ou votre adresse e-mail.
- `mail_contact` : E-mail à indiquer en tant qu'expéditeur des courriers envoyés par SFerConfig
- `mail_foot` : Message à ajouter à la fin des courriers envoyés par SFerConfig
- `sfercfg_base_url` : URL de SFerConfig
- `db_host` : machine contenant la base de données utilisée par SFer
- `db_login` : login pour l'accès MySQL
- `db_password` : mot de passe pour l'accès MySQL
- `db_name` : nom de la base de données utilisée par SFer. Ce nom a été défini précédemment dans `~/sfer/sfer_config.xml`

3 Utilisation

Une fois l'installation et la configuration des différents composants réalisée, le plus dur du travail a été fait. Cette partie décrit comment utiliser chaque programme et mettre en pratique ces nombreux réglages définis dans la partie précédente.

3.1 web2sql

3.1.1 Exécution

web2sql se charge de lire chaque forum de discussion, suivant les paramètres définis dans son fichier de configuration, et de les stocker dans la base de données définie elle aussi dans ce même fichier.

Pour ce faire, il suffit de lancer son exécution :

```
$ web2sql
```

web2sql affichant sur la sortie d'erreur toute information sur le déroulement des téléchargements, il peut être utile de récupérer ces informations dans un fichier log, de sorte à pouvoir l'analyser par la suite si besoin :

```
$ web2sql 2> mon_fichier_log.txt
```

3.1.2 Informations en sortie et fonctionnement

Un début d'analyse se déroulant correctement devrait prendre la forme suivante :

```
web2sql 1.0beta  
Written by Fremex <fremex@planet-shitfliez.org>
```

```
This is free software; see the source for copying conditions. There is NO  
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

3.2. SFER

```
Forum      1 : PC
-----
Connecting to MySQL database forum_1@localhost:3306...Ok
forum 1 page 1 (25 matches)
forum 1 page 2 (25 matches)
forum 1 page 3 (25 matches)
forum 1 page 4 (25 matches)
forum 1 page 5 (25 matches)
forum 1 topic 1/89 9772940 page 1175 (14 matches)
forum 1 topic 1/89 9772940 page 1176 (0 matches)
forum 1 topic 2/89 10461445 page 2 (0 matches)
forum 1 topic 3/89 10957128 page 1 (4 matches)
forum 1 topic 3/89 10957128 page 2 (0 matches)
```

Comme on peut le voir, web2sql se connecte à la base de données `forum_1` du serveur `localhost`, parcourt les 5 premières pages listant les topics du forum `PC` et récupère à la fin de chaque topic les messages qu'il n'a pas encore en base de données. Pour chaque topic, lorsqu'il ne trouve plus rien (0 matches), il en conclut qu'il a atteint la dernière page du topic et passe au topic suivant.

Remarque : si web2sql doit être interrompu, il est préférable de le faire dans les phases de parcours des pages listant les topics (par exemple `forum 1 page 3 (25 matches)` dans l'exemple ci-dessus). Dans le cas contraire, web2sql tentera d'annuler la lecture des messages du forum en cours, mais malgré les tests, cette fonctionnalité peut présenter des défauts.

3.2 SFer

3.2.1 Exécution

SFer s'exécute après web2sql, dans le cas où l'on souhaite utiliser ce dernier. Il récupère les nouveaux messages présents dans les bases de données choisies dans `input_mysql_config.xml`, les analyse, et génère un ensemble de pages HTML grâce au plugin de sortie.

Son lancement se fait ainsi :

```
$ sfer2
```

De la même façon que web2sql, l'ensemble des messages d'information sont envoyés sur la sortie d'erreur, qui peut éventuellement être redirigée vers un fichier log :

```
$ sfer2 2> mon_fichier_log.txt
```

3.2. SFER

3.2.2 Paramètres

SFer dispose également d'un ensemble de paramètres qui sont à votre disposition. La liste peut être obtenue à l'aide du paramètre `-h`.

```
$ sfer2 -h
SFer 2.0beta
Written by Fremem <fremem@planet-shitfliez.org>
```

```
This is free software; see the source for copying conditions.
There is NO warranty;
not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

```
sfer2 [options]
```

```
-h : display this help message
```

```
-p : parse only mode
```

```
-s : statistic creation only mode
```

```
-a : force stats creation, even for unmodified forums
```

```
-b id|list : begin parse/statistic from forum 'id'
              (can be useful after an abnormal termination)
```

```
-f id|list : work only on forum 'id' or display forum 'list'
```

```
-c : don't try to create sub databases (for db_split mode)
```

```
-n : create new main database if not exists (first use)
```

```
-e id|list|all : empty forum 'id', 'all' forums or display forum 'list'
```

```
-k id|list|all : delete forum 'id', 'all' forums or display forum 'list'
                  delete = empty + remove forum from forum list
```

```
-d dir : configuration directory for sfer and plugins. Default : $HOME/.sfer
```

```
-y : turn off interactif prompting
```

Voici une description plus détaillée de ces paramètres :

- `-p` : SFer se contentera d'analyser les nouveaux messages mais ne générera pas de statistiques
- `-s` : SFer se contentera de générer des statistiques, sans analyser les nouveaux messages
- `-a` : SFer générera des statistiques pour tous les forums, même ceux qui n'ont pas été modifiés
- `-b` : L'analyse des messages et/ou la génération des statistiques commencera à partir du forum dont l'identifiant a été fourni en paramètre. Tous les forums précédents seront ignorés.

3.2. SFER

- `-f` : L'analyse des messages et/ou la génération des statistiques ne se fera que pour le forum dont l'identifiant a été fourni en paramètre.
- `-c` : Si le mode `db_split` est actif, force SFer à ne pas essayer de créer de sous-bases de données. Ce paramètre permet de gagner du temps quand il n'y a pas eu de nouveaux forums depuis la dernière exécution.
- `-n` : Demande à SFer de créer sa base de données principale, si elle n'existe pas déjà
- `-e` : Vide les informations concernant le forum dont l'identifiant a été fourni en paramètre. Toutes les données résultant de l'analyse de ce forum seront effacées. A l'exécution suivante, SFer relancera de nouveau l'analyse de chaque message de ce forum.
- `-k` : Même chose que `-e`, et supprime définitivement le forum correspondant de la base de données de SFer. Il faudra de nouveau ajouter le forum (avec `SFerConfig`) pour pouvoir de nouveau faire des statistiques de celui-ci.
- `-d` : Répertoire où SFer doit aller chercher ses fichiers de configuration, y compris les plugins. Ainsi, il est possible de gérer plusieurs bases de données distinctes, chacune contenant ses propres forums.
- `-y` : Répond "oui" à toutes les demandes de confirmation.

3.2.3 Informations en sortie et fonctionnement

3.2.3.1 Analyse des messages

SFer fonctionne en deux phases. Tout d'abord, il va analyser les nouveaux messages pour chaque forum, sauf si un paramètre va dans le sens contraire.

L'analyse de chaque forum prend un forme similaire à :

```
Site 1 Forum "PC" (id=26 num=1)
2966164 posts found in database
Parsing new messages : 7122 messages parsed.
```

Dans cet exemple, SFer vient de traiter le forum PC, qu'il identifie sous le nombre 26 et reconnu sous l'identifiant 1 sur son site d'origine. Ce forum comptait 2966164 messages déjà analysés et 7122 nouveaux viennent de l'être.

Lorsque le nombre de messages dépasse 10000, SFer affiche par intervalle de 10000 messages le temps de calcul déjà réalisé, de sorte à montrer l'avancement. Certains forums pouvant compter plusieurs millions de messages, il est important de savoir si l'analyse se déroule correctement.

3.2.3.2 Génération des statistiques

La deuxième phase de SFer est la génération des statistiques. Pour reprendre l'exemple du forum PC, voici une partie de la sortie obtenue :

3.3. SFERCONFIG

```
|--Forum PC (708 sec)
  |--Topic Mon buro c'est le plus boooooo (473 sec)
  |--Topic Ogame.fr vous connaissez ? (351 sec)
  |--Topic les stats du forum PC ! ! ! ! (346 sec)
  |--Topic [BlaBla - FORUM PC] (389 sec)
  |--Topic La 9600 pro suit-elle encore ? (342 sec)
  |--Topic image de vos configure 'in game' (345 sec)
  |--Topic Notpron: le jeu le plus dur du net :- ) (202 sec)
```

Dans cet exemple, la page de statistiques du forum PC a été générée en 708 secondes. Certains topics ayant une taille suffisamment importante (définie dans SFerConfig), des statistiques sont également générées. Un lien vers chacune d'entre elles se trouvera sur la page de statistiques du forum.

3.3 SFerConfig

SFerConfig est une interface PHP permettant de paramétrer les comptes utilisateur et de réaliser la plupart des opérations d'administration de SFer. Comme une illustration vaut mieux que de longs discours, un aperçu de l'interface est représenté sur la figure 3.1 page 22.

Remarque importante : par défaut, un compte administrateur est disponible (login : admin, mot de passe : admin) et son mot de passe doit impérativement être modifié pour éviter tout problème de sécurité.

De façon générale, SFerConfig se découpe trois parties : utilisateur, modération et administrateur. Cette section propose une description de chacune d'entre elles.

3.3.1 Partie utilisateur

Une fois inscrit depuis la page d'accueil de SFerConfig, l'utilisateur se voit proposer quatre fonctionnalités principales.

- **Modification du profil** : permet de changer de pseudonyme, de mot de passe et de régler ses informations et sa carte de visite. Ces données seront récupérées par SFer et ajoutées dans les statistiques.
- **Gestion des pseudonymes** : un utilisateur est autorisé à ajouter un ou plusieurs pseudonymes à son compte. Les pseudonymes autorisés sont ceux qui ont été récupérés par SFer lors de l'analyse des messages. Ainsi, si un utilisateur ajoute un ou plusieurs pseudonymes à son compte, ils seront fusionnés et apparaîtront sous la forme d'un seul pseudonyme dans les statistiques. Cette fusion est réversible.
- **Présence dans les classements** : si un utilisateur souhaite que l'un de ses pseudonymes ne soit plus pris en compte dans les statistiques, il lui suffit d'utiliser cette option.



FIG. 3.1 – Capture d'écran de SFerConfig

- **Ancienneté dans les classements** : si un utilisateur souhaite que l'ancienneté¹ d'un de ses pseudonymes ne soit plus pris en compte dans les statistiques, il lui suffit d'utiliser cette option. Cela peut s'avérer utile sur des forums où d'anciens pseudonymes ont été libérés et où de nouveaux utilisateurs se les sont appropriés.

3.3.2 Partie modération

Si un utilisateur a été promu au rang de modérateur par l'administrateur, il a accès aux options de modération, permettant globalement de faire les mêmes manipulations que l'utilisateur sur n'importe quel compte.

- **Gestion des bannissements** : permet de bannir un ou plusieurs pseudonymes. Ces pseudonymes resteront dans la base de données mais ne seront plus pris en compte lors du calcul des statistiques.
- **Gestion de l'ancienneté** : permet de désactiver l'ancienneté d'un ou plusieurs pseudonymes. Ces pseudonymes resteront intacts dans la base de données mais leur ancienneté ne sera plus prise en compte lors du calcul des statistiques.
- **Gestion des regroupements** : permet de modifier des regroupements de pseudonymes
- **Bannissement de lien** : cette fonctionnalité sert à ne plus prendre en compte certains liens dans les statistiques.

3.3.3 Partie administration

La partie administration est réservée à l'administrateur et permet de régler la plupart des paramètres réutilisés par SFer.

- **Ajout/suppression de forum** : cette section sert à ajouter des forums qui seront ensuite pris en compte par SFer. Les forums ajoutés doivent être décrits exactement de la même manière dans le fichier de configuration du plugin d'entrée `input_mysql_config.xml`.
- **Définition du site** : définition du site regroupant les forums dont SFer génère les statistiques.
- **Options des statistiques** : cette section permet de paramétrer les statistiques de l'ensemble des forums concernés. Il est également possible de spécifier des valeurs particulières pour chaque forum.
- **Niveau des utilisateurs** : cette fonctionnalité sert à attribuer un niveau à un utilisateur inscrit.
- **Smileys** : permet de définir les différents "code de smiley" qui seront interprétés de façon particulière par SFer. Il est possible de fournir une URL vers l'image correspondant à chaque smiley ; cette URL sera reprise lors de la génération des statistiques. Les codes de smileys ne seront pris en compte que pour les messages qui n'ont pas déjà été analysés.

¹la date et heure du premier message

4 A propos

4.1 Licence

SFer, incluant l'ensemble de ses composants, est distribué sous licence GPL. Elle garantit les quatre libertés suivantes :

1. la liberté d'exécuter le logiciel, pour n'importe quel usage
2. la liberté d'étudier le fonctionnement d'un programme et de l'adapter à ses besoins, ce qui passe par l'accès aux codes sources
3. la liberté de redistribuer des copies
4. la liberté d'améliorer le programme et de rendre publiques les modifications afin que l'ensemble de la communauté en bénéficie

Une copie de la licence se trouve dans le fichier COPYING fourni avec les sources de chaque composant.

4.2 Contacter l'auteur

L'auteur de SFer peut être contacté par mail à l'adresse suivante :
`fremen _AT_ planet-shitfliez.org`

Tout encouragement ou critique sont les bienvenus, une réponse sera retournée dans la mesure du possible.