

Sujet 17 : Exploitation de données financières de Google Finance en C#

Introduction :

Il convient tout d'abord de signaler que le titre de ce projet n'est pas en adéquation avec son contenu. Nous ne nous occupons pas de l'exploitation mais de la récupération des données de Google Finance. Ce décalage provient du fait que nous avons récupéré un projet existant. La méthode de récupération de données utilisée l'année dernière s'est avérée inefficace, nous avons donc décidé, avec l'accord de notre responsable de projet, de recommencer un nouveau projet en utilisant une méthode de récupération plus fiable.

Notre projet de spécialité se compose de deux parties distinctes. La première partie concerne l'implémentation d'une DLL en C#. Cette DLL permet au programmeur une récupération rapide et efficace des données de Google Finance. Elle sera notamment utilisée par les futurs 3A dans le cadre des projets d'exploitation de données financières. La deuxième partie du projet consiste en l'implémentation d'un Windows Form. Le Windows Form se sert de la DLL pour proposer à l'utilisateur une récupération personnalisée des données financières de Google.

Première Etape : Récupération du projet de l'an dernier et prise de décision

En début de projet, nous avons récupéré les programmes réalisés l'an dernier. Nous étions déjà confrontés à un premier choix : continuer et corriger le projet récupéré ou repartir de zéro avec une nouvelle méthode de récupération des données financières.

Durant les premiers jours, nous avons essayé de comprendre au mieux le code qui nous était fourni, afin de cerner quelle partie de la récupération de données posait problème. En parallèle, nous cherchions de notre côté une méthode simple et efficace pour obtenir ces données. Nous en avons finalement trouvé une : on effectue une requête http sur le site de Google. Cette requête nous rend un fichier au format csv, qui contient justement les données financières attendues. Suite à cette découverte, nous avons décidé de commencer un nouveau projet sans se servir des programmes fournis.

Deuxième Etape : Elaboration de la DLL

Une fois notre méthode de récupération trouvée, nous avons pu commencer l'élaboration de notre DLL. Elle s'appelle GoogleFiMag et comporte deux classes principales.

La première est la classe DataFI. C'est la classe que l'utilisateur de la DLL doit instancier pour récupérer ses données financières dans une matrice à trois dimensions. La première dimension représente les indices boursiers, la deuxième les dates, et la troisième correspond aux champs observés (par exemple : Open, Low, Close ...).

La deuxième classe principale est GoogleFI. C'est typiquement la classe que l'utilisateur doit instancier pour gérer l'exportation de ses données. Conformément aux attentes de notre responsable de projet, nous proposons plusieurs formats d'exportations possibles. Dans tout les exemples d'exportations qui suivent, nous récupérerons les données 'Low' et 'Close' sur l'indice Google entre le 14 et le 17 Février 2011.

→ **Exportation en Csv** : le fichier s'ouvre avec excel et se présente sous la forme suivante :

	A	B	C	D
1	Date	Low	Close	
2	14-Feb-11	620.04	628.15	
3	15-Feb-11	623.1	624.15	
4	16-Feb-11	622.1	624.22	
5	17-Feb-11	620.28	625.26	
6				

→ **Exportation en Xml** : on peut ouvrir le fichier avec un navigateur web. Les données sont récupérées avec la syntaxe suivante :

```
<?xml version="1.0" standalone="yes" ?>
- <CSVRoot>
- <CSVTable>
  <Date>14-Feb-11</Date>
  <Low>620.04</Low>
  <Close>628.15</Close>
</CSVTable>
- <CSVTable>
  <Date>15-Feb-11</Date>
  <Low>623.1</Low>
  <Close>624.15</Close>
</CSVTable>
- <CSVTable>
  <Date>16-Feb-11</Date>
  <Low>622.1</Low>
  <Close>624.22</Close>
</CSVTable>
- <CSVTable>
  <Date>17-Feb-11</Date>
  <Low>620.28</Low>
  <Close>625.26</Close>
</CSVTable>
</CSVRoot>
```

- **Exportation en Mdf** : pour réaliser cette exportation, on a attaché à notre DLL une base de données sous la forme d'un fichier mdf. Cette base comporte une seule table appelée historique, qui est elle-même composée des cinq colonnes PriceOpen, High, Low, PriceClose et Volume. Lorsque l'utilisateur demande une exportation en mdf, la base de données est remplie puis copiée là où l'utilisateur le demande. Le contenu du fichier mdf ainsi généré peut se lire avec Visual Studio de la manière suivante :

	Indice	Date	PriceOpen	High	Low	PriceClose	Volume
▶	GOOG	14/02/2011 00:0...	NULL	NULL	620,04	628,15	NULL
	GOOG	15/02/2011 00:0...	NULL	NULL	623,1	624,15	NULL
	GOOG	16/02/2011 00:0...	NULL	NULL	622,1	624,22	NULL
	GOOG	17/02/2011 00:0...	NULL	NULL	620,28	625,26	NULL
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL

- **Exportation en base de données** : Si l'utilisateur choisit ce mode, il doit créer par lui-même sa base de données (par exemple en utilisant Sql Server Management Studio 2008). Il doit alors donner à la DLL les paramètres de sa base de données, c'est-à-dire le nom du server (localhost\SQLEXPRESS), le nom de sa base et le nom de la table. L'exportation en base de données crée la table dans sa base et la remplit avec les valeurs financières attendues. Un exemple de visionnage des résultats est donné ci-dessous.

	Indice	Date	Low	PriceClose
1	GOOG	2011-02-14 00:00:00.000	620,04	628,15
2	GOOG	2011-02-15 00:00:00.000	623,1	624,15
3	GOOG	2011-02-16 00:00:00.000	622,1	624,22
4	GOOG	2011-02-17 00:00:00.000	620,28	625,26

Outre les deux classes principales, la DLL comporte aussi les classes suivantes. La classe Requete.cs sert à gérer les requêtes http envoyées sur Google. Parser.cs permet de lire les fichiers csv rendus par Google à la suite des requêtes, et ainsi de récupérer les données financières téléchargées. ParseCsvToXml.cs permet de transformer un fichier Csv en Xml (nécessaire pour une exportation en Xml) et Parserxml.cs permet de lire des fichiers Xml. Cette dernière classe a été ajoutée pour permettre une utilisation simplifiée du Windows Form que nous allons voir maintenant.

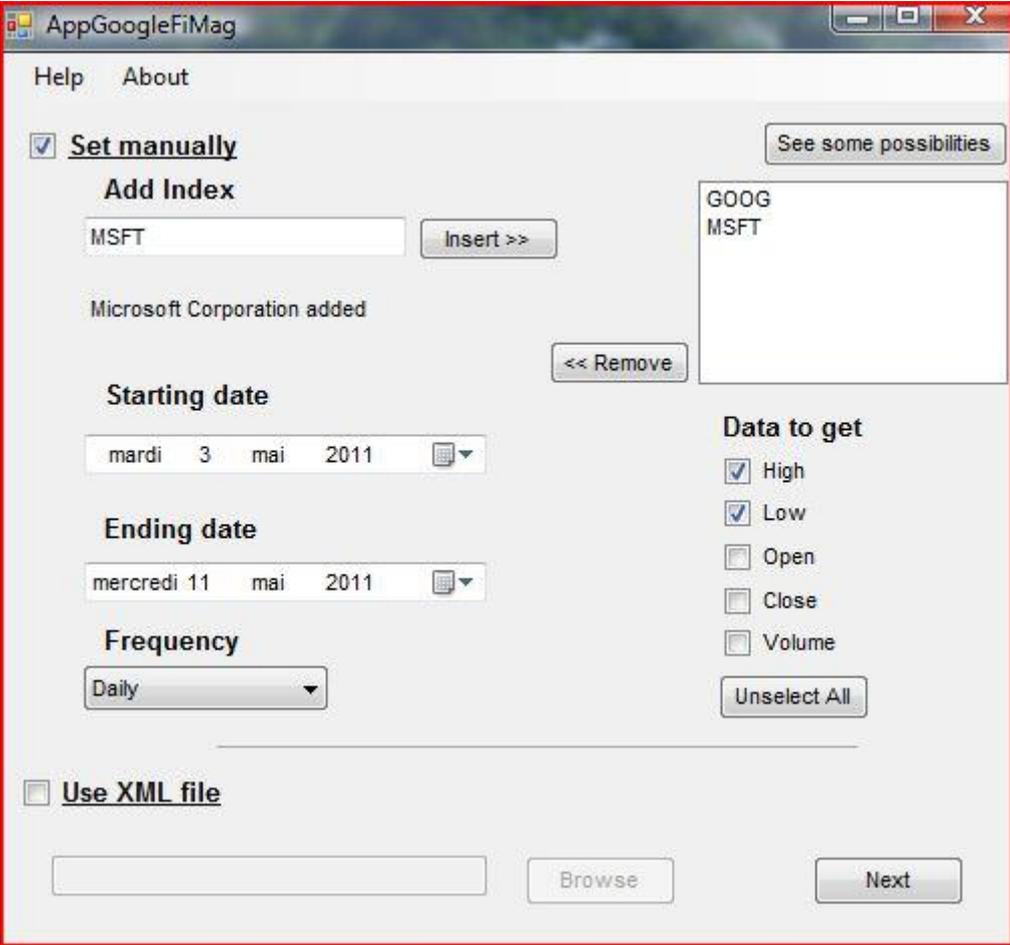
Troisième Etape : Conception du Windows Form

Le Windows Form permet de récupérer des données financières sans passer par la programmation. L'utilisateur dispose d'une interface graphique simple d'utilisation qui le

guide dans ses choix de données et d'exportation. Le Windows Form se compose de deux fenêtres distinctes.

La première fenêtre demande à l'utilisateur de préciser la manière dont il veut entrer ses paramètres. Il peut soit les entrer manuellement, soit dans un fichier Xml. S'il choisit le mode manuel, il doit alors préciser la date de départ et de fin de la récupération, ainsi que la liste des indices boursiers à observer et la fréquence d'affichage. S'il choisit le mode Xml, il doit entrer tout ses paramètres dans un fichier Xml selon une syntaxe particulière qui a été défini avec l'équipe travaillant sur le projet Yahoo Finance.

Fenêtre 1 : Réglage des paramètres

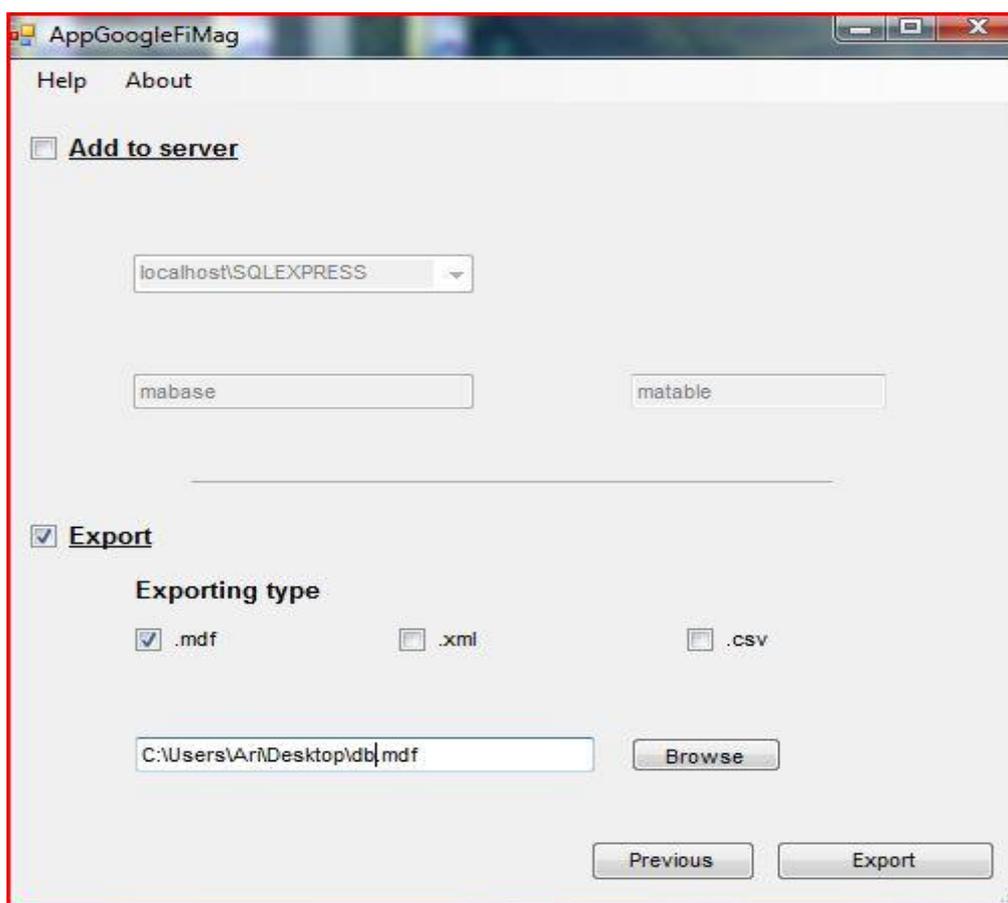


The screenshot shows the 'AppGoogleFiMag' application window. At the top, there are 'Help' and 'About' menu items. The main interface is divided into several sections:

- Set manually:** A checked checkbox. To its right is a button labeled 'See some possibilities'.
- Add Index:** A text input field containing 'MSFT' and an 'Insert >>' button. Below it, the text 'Microsoft Corporation added' is displayed.
- Starting date:** A date picker showing 'mardi 3 mai 2011'.
- Ending date:** A date picker showing 'mercredi 11 mai 2011'.
- Frequency:** A dropdown menu currently set to 'Daily'.
- Data to get:** A list of checkboxes: 'High' (checked), 'Low' (checked), 'Open' (unchecked), 'Close' (unchecked), and 'Volume' (unchecked). Below this list is an 'Unselect All' button.
- Remove:** A '<< Remove' button is located between the 'Add Index' and 'Data to get' sections.
- Index List:** A list box on the right contains 'GOOG' and 'MSFT'.
- Use XML file:** An unchecked checkbox at the bottom left.
- File Selection:** An empty text input field, a 'Browse' button, and a 'Next' button are located at the bottom right.

La deuxième fenêtre sert à préciser le type d'exportation souhaité. La case 'Add to server' concerne le cas d'une exportation en base de données. L'utilisateur entre sa base de données avec son serveur et son nom de table et celle-ci sera remplie avec les données attendues en cliquant sur le bouton 'Add Server'. La case 'Export' permet de choisir une ou plusieurs exportations dans les formats Mdf, Xml et Csv. Il faut ensuite choisir un dossier à l'aide du bouton 'Browse' pour conserver les fichiers générés.

Fenêtre 2 : Choix de l'exportation



En ce qui concerne l'importation des fichiers xml, la syntaxe définie avec le groupe Yahoo Finance est la suivante (voir image ci-dessous) :

- Entre les balises <Index>, il faut écrire les indices boursiers dont on veut récupérer les données financières.
- Entre les balises <Param>, il faut écrire uniquement les champs que l'on veut observer parmi les cinq proposés.
- Entre les balises <DateDebut> et <DateFin>, il faut écrire les dates d'observation initiale et finale au format anglais.
- Entre les balises <Frequence>, il faut préciser si vous voulez un affichage quotidien ou hebdomadaire. Si vous écrivez autre chose que daily ou weekly, la fréquence daily sera retenue.
- Entre les balises <ExportParam>, il faut écrire une seule balise parmi <BDD>, <Csv>, <Xml> et <Mdf>. Si vous en écrivez plusieurs, un message d'erreur explicite vous sera proposé lors de l'exportation. Entre vos balises d'exportation, il faut écrire l'adresse absolue du fichier exporté et sans son extension.

Syntaxe Xml pour définir les modalités de la récupération

```
<?xml version="1.0" encoding="UTF-8"?>
<RecupData>
  <ListeIndex>
    <Index> Index 1 </Index>
    <Index> Index 2 </Index>
    <Index> ... </Index>
  </ListeIndex>
  <Champs>
    <Param> Open </Param>
    <Param> High </Param>
    <Param> Low </Param>
    <Param> Close </Param>
    <Param> Volume </Param>
  </Champs>
  <DateDebut> Date au format anglais (2011-05-21) </DateDebut>
  <DateFin> Date au format anglais (2011-05-21) </DateFin>
  <Frequence> Ecrire daily ou weekly </Frequence>
  <ExportParam>
    <BDD>
      <Adresse> Adresse de la base de données </Adresse>
      <Login> Nom de l'utilisateur </Login>
      <Password> Mot de passe </Password>
      <Port> Port </Port>
      <NomBDD> Nom de la base </NomBDD>
      <NomTable> Nom de la table </NomTable>
    </BDD>
    <Csv> Adresse du fichier exporté </Csv>
    <Xml> Adresse du fichier exporté </Xml>
    <Mdf> Adresse du fichier exporté </Mdf>
  </ExportParam>
</RecupData>
```

Quatrième Etape : Documentation et ajouts de fonctionnalités

Nous avons conscience que notre projet peut être utilisé à la fois par les 3A de la filière finance comme par les futurs 2A dans le cadre d'un projet de spécialité. Il est donc primordial de rédiger une documentation claire pour faciliter l'utilisation de la DLL et du Windows Form.

D'une part, nous avons ainsi réalisé une documentation de notre dll à l'aide de l'outil Sandcastle. Les classes que l'utilisateur ne peut instancier (Requete.cs, Parser.cs etc...) ne sont pas visibles dans la documentation puisque celui-ci n'aura pas à s'en servir. En cliquant sur les différentes classes, l'utilisateur accède aux explications sur leurs méthodes.

D'autre part, nous avons rédigé un manuel utilisateur pour expliquer le bon fonctionnement du Windows Form. Contrairement à la documentation de la DLL qui s'adresse au programmeur, le manuel s'adresse à toute personne voulant récupérer les données financières de Google sans passer par la programmation. Nous présentons ici sa table des matières.

Documentation DLL : Page d'accueil

Classes

	Class	Description
	ConnectivityException	Exception to handle the connectivity, you have to but connected to use this DLL
	DataFI	Class which allow you to stock the historical datas that you retrieve from the internet. There is a lot of fonction that allow you to get what you need after you created the object itself. This class is mainly about the matrix mat_valeurs with every data in it
	GoogleFI	This classes allows you to use all of the export methods of the GoogleFIMag dll. In order to instance this classes, you have to indicate the starting date and the ending date of the period for recovering data from Google Finance. You also have to specify if you want daily or weekly data. Finally, you have to indicate the list of indexes you want infos on. GoogleFI class provides a lot of export methods in different formats such as XML, CSV, MDF or double matrix. It is possible to obtains information about every official Reuter indexes, since it's recorded on Google Finance server. Datas are "Open" : open price for a specific day "Close" : closing price for an index. "Low" and "High" for a day and an index, and "Volume" which is the number of transactions for an index on a day.
	historique	
	linkDataContext	
	UnknownIndexException	Exception thrown when the index is unknown on GOOGLE FINANCE

Manuel utilisateur du Windows Form

Table des matières

1	Première fenêtre	3
1.1	Configurer la recherche manuellement	4
1.2	Configurer la recherche à l'aide d'un fichier .xml	6
2	Deuxième fenêtre	8
2.1	Exportation dans une base de données	10
2.2	Exportation en fichiers	11
2.2.1	Exportation en .mdf	12
2.2.2	Exportation en .xml	12
2.2.3	Exportation en .csv	12
3	Les messages d'avertissement	13
4	Norme pour le fichier XML	14
5	Structure de la table de la base de données	15

Enfin, nous avons décidé d'ajouter une fonctionnalité à notre Windows Form pour permettre une utilisation simplifiée à l'utilisateur. Google Finance ne répertorie pas nécessairement tous les indices boursiers du marché, pourtant il est indispensable de connaître à l'avance quels sont les indices sur lesquels on va effectivement pouvoir récupérer des données. Pour résoudre ce souci, nous avons décidé de tester toutes les chaînes de caractères ayant entre une et quatre lettres et de regarder si elles correspondaient à un symbole d'un indice proposé par Google. Nous avons regroupé les résultats dans une base de données, ainsi nous connaissons à l'avance tous les indices boursiers que Google Finance répertorie. Pour y avoir accès, l'utilisateur doit ouvrir le Windows Form et cliquer sur 'See Some Possibilities'. On obtient alors le tableau ci-dessous.

Listes des indices proposés par Google

The screenshot shows a Windows application window titled "AppGoogleFiMag" with a standard Windows title bar (minimize, maximize, close buttons). Inside the window, there is a dialog box titled "Select what you want to insert". The dialog box contains a table with four columns: "Symbol", "Name", and "Exchange". The first row is highlighted in blue. Below the table, there are two buttons: "Insert" and "Ok".

Symbol	Name	Exchange
AAAA	AAA Public Adjusting Group Inc	OTC
AAAE	AAA Energy Inc	PINK
AAIF	Aberdeen Asian Income Fund Ltd.	LON
ABWG	A.B. Watley Group Inc.	PINK
AUO	AU Optronics Corp. (ADR)	NYSE
PKM	Preferred Plus Trust Ser ALL 1 Tr Cf CL A Pfd Pr...	NYSE
PKN	Powershares Glob Nuc Energy (ETF)	NYSE
PKOH	Park-Ohio Holdings Corp.	NASDAQ
PKO	PIMCO Income Opportunity Fund	NYSE
PKPL	Park Place Energy Corp	PINK
SIX	Six Flags Entertainment Corp	NYSE
PKTX	ProtoKinet	OTC
PKT	Procera Networ	AMEX
PKWY	Parkway Bank	OTC
PKW	PowerShare Buyback Achievers Fund (ETF)	NYSE
AURI	Wellstone Filter Scienc	OTC
PKX	POSCO (ADR)	NYSE

Conclusion :

Nous avons tous les trois beaucoup apprécié ce projet de spécialité. En premier lieu, il présente un intérêt majeur : récupérer des données financières est la première étape à effectuer

avant de réaliser n'importe quelle analyse financière. En particulier, nous espérons que les 3A utiliseront notre logiciel pour les projets d'exploitation de données financières.

En second lieu, ce projet de spécialité nous a offert une excellente initiation au C#, langage fortement utilisé en finance, ainsi qu'en conception de Windows Form, gestion de fichiers mdf ou exportations sous différents formats. Nous avons également découvert qu'il était possible de récupérer des données financières sur Google Finance à partir d'une simple requête http, ce qui constitue d'ailleurs la base de notre projet.

Enfin, nous nous sommes très bien entendus au sein du groupe. Nous avons décidé de travailler tous ensemble à l'Ensimag pour s'entraider plus facilement et pour communiquer directement. Dans une atmosphère conviviale, nous travaillions efficacement et pouvions contrôler l'avancement de chacun. Nous garderons un très bon souvenir de ce projet de spécialité.