# SemSync

The reason for me to start and continue this project is not to build up a super user friendly end user product, but more to have some playground to experiment with. You will also find some things even I’m not so sure that they are good practice: This project will make use of the “var” keyword, just to see, if it will do harm to the readability of the code.

## What’s the goal?

SemSync is a project for synchronizing entities between different object stores. Currently this is a ***project***, not a product. If you trust in my capabilities to write nice programs and want to try synchronizing without installing Visual Studio, then download the file “**Sync Outlook with Xing**” – this is a compiled version with a very simple user interface. The current implementation is done for synchronizing contact entities and comes with “Connectors” to

1. the file system (as one big or many small xml files, plus one implementation using vCards)
2. Microsoft Outlook 2007
3. Microsoft Outlook 2003 (also compatible with Outlook 2007)
4. Xing contacts using web scraping technology
5. A simple WCF online store

***Please read this document carefully,   
because this project may alter your native data under unexpected conditions!***

## What’s in the package?

The project consists of a base library which contains the entity, helper classes and the execution engine. The engine will execute command which contain parameters and up to three “connectors” (source, target and baseline). You can think of a command as processing data streaming from one connector (source) to another (target). Some commands also involve a third data stream (baseline), like the merge command which can detect changes of a source and a target stream by comparing both to the baseline.

The internal data representation is a proprietary class that is serializable. The class might change (because the current implementation does not follow xNL/xAL) in the future, but currently there are more attractive goals for working with this project.

## The connectors

Connectors do read/write from/to data sources like file system, online storage and processes like Microsoft Outlook. If a connector is useful for you depends strongly on your expectations: some connectors do only read from a source, because that source is does not accept writing (like most social networking sites) – some sources do provide only very few information. The source “Facebook” does not support reading much data, but you might be able to extract useful pictures from there.

### The File System connectors

The file system connectors (with the exception of the vCard implementation) do serialize the internal objects or a list of these objects to the file system. The vCard implementation of the file system connector does currently not support all properties of the internal Contact class.

### The Online Storage connector

This connector communicates with a WCF service that is also part of the package. You will have to modify the service and the connector to match your needs (security, physical storage etc.).

### The Xing connector

Xing is an online business community to handle contacts.

You can configure the login credentials for this portal inside the app.config. If there is no password configured, a login dialog will appear to ask for login credentials. The user id can be preconfigured inside the app.config, even if you don’t want to store the password locally. The user id does always have the focus, even if there is one configured.

The connector logs in, detects the contacts from the own contact list and downloads the vCards from the Xing portal. It then converts the vCards into the internal contact representation. You can combine the Xing connector and the file system connector to export the Xing contacts into the file system.

There is a caching mode for Xing that will download the information and use that downloaded information in subsequent executions. This mode is for debugging purpose only and should not be activated by a normal user or to synchronize real data.

### The Facebook connector

The Facebook api does not provide much information that can be handled in a contact-application. So this connector is more for completion and photo-extraction.

## Working with contacts

The internal contact class does have an Id to identify the contact. This Id is a Guid and will be generated when creating an instance of the class. The engine is capable to match the contacts using this Id. If you do export contacts from different sources, you can also match the contacts using the name and replace the Id in the target (the target connector will read the contacts, the Ids will be overwritten and the target connector will then save the contacts).

Some connectors cannot write contacts (like Xing, because Xing does not allow altering the contact information) – other connectors **may change the external data even when only reading**: the Outlook connector writes back the Id into Outlook as a *user defined property* to have always the same Id for an exported contact.

## Importing data

I’m currently using the following sequence to import and synchronize contacts from Xing to Microsoft Outlook:

1. Delete previous export/work files
   * Otherwise the contacts exported last time will stay in place while adding the current contacts from Xing
2. Export from Outlook to file system
   * This will create a file system representation of the Outlook contacts. Using the file system representation is much better to handle for the connectors.
3. Export from Xing to file system
   * This will download the contacts from Xing. You might configure the credentials in the app.config file to bypass the login screen. You also might use the IE cookies if you did save the login while using the IE. The download will be done using the normal Html UI of Xing – this may be a problem when Xing does change the GUI.
4. Normalize Xing(fs)
   * Some people do use other spellings of the company name than others (while working for the same company). Normalizing does allow you to specify replacements for strings, so that all contacts do use the same “wording”.
5. Normalize Outlook(fs)
   * Same as the one before – just for the outlook contacts.
6. Match Outlook to Xing by name
   * This will do lookups for the Xing contacts inside the Outlook contacts. We need something like this, because Xing is not capable to write back the Sync-Id.
7. Add missing from Outlook(fs) to Xing(fs)
   * For comparison in third party tools it’s much easier if we have no missing contacts.
8. Add missing from Xing(fs) to Outlook(fs)
   * For comparison in third party tools it’s much easier if we have no missing contacts.
9. Merge high evidence from Xing(fs) to Outlook(fs)
   * Some of the information can be overwritten without user intervention. E.g. if the information is only available on one side or the other side is out of scope (date of birth > 01.01.2020)
10. Merge high evidence from Outlook(fs) to Xing(fs)
    * Same as above but other direction.
11. Detect Conflicts
    * This will include the user by presenting a dialog to “merge” the conflicting information. The user should select the correct information (color it green).
12. Merge Xing to Outlook via BeyondCompare
    * This will open a third party tool (BeyondCompare) to show still not matching information.
13. Ask if we want to overwrite Outlook
    * Before writing to Outlook we will ask the user if she is really sure … ;-)
14. Import all from file system to Outlook
    * Now we write all the data from the manipulated Outlook file system representation to the Outlook store.

After this sequence Outlook “should” contain all new and updated information. If that’s not the case you might file a bug and provide enough information to let me reproduce the problem in detail and on my person PC while debugging the program.

## Sample scripts

### Sync with Xing

This script does include commands to sync contacts from Xing to Outlook. This script does use an external tool (BeyondCompare) to solve merge conflicts. You will need to download the tool if you want to use this script. Actions:

1. Cleanup working folder
2. Export Outlook contacts
3. Export Xing contacts
4. Merge automatically missing entries in both files
5. Open external tool to perform manual merge fixing
6. Import into Outlook
7. Remove duplicate calendar entries from Outlook (excluded in sample script)

You might include the last entry of the script. I’ve tested that process many times with my own appointments, but I cannot guarantee that it will not remove non-duplicate entries, so I’ve commented it out. The reason to include this is that Outlook automatically adds calendar entries for contact birthdays – so when importing from many sources, you might blow up your calendar.

### Sync with WCF (Simple)

This is very similar to “Sync with Xing” with the exception that it will also upload the synchronized list to the Online Storage WCF service (in contrast to Xing, that one is capable to update data).

### Sync TESTS

**This is my experimental script and will include changing commands. Review it carefully before executing.**

## Authoring connectors

Writing a new connector is very simple: just inherit from Sem.Sync.SyncBase.StdClient and override the two abstract methods. The complexity might come with implementing these two methods, because you will need to provide a conversion from your “native” data to the StdContact class. Depending on the type of data you read/write this might be easy or complex.

The connectors are instantiated by a class factory. In the scripts just use the full qualified class name (include the assembly name if that is different to the namespace) and the engine will do the rest.

## Sync Outlook with Xing

This zip file includes a very simple user interface to download contacts from Xing and synchronize with Microsoft Outlook. Unfortunately I didn’t finish the user interface to compare and merge the contacts, so you need a third party tool: BeyondCompare.

When starting the executable in that zip, you will be presented a disclaimer that will tell you that I will not be responsible for any damage this software will do to your PC and that you understand that you are responsible for performing a backup of your data, not me. I don’t want to get angry emails of people who download software for free that I have written in hours of work and that lost a space in a street name (or even the whole address book), so backup your data before executing software that is designed to alter your data.

If you tell the software that you do understand the terms, you will be presented a very spartanic user interface with exactly one button. After downloading the Xing-contacts (you will need to enter your credentials) and exporting the Microsoft Outlook contacts, the program will use its own UI to let you merge conflicts and then ask if you want to import the result. Solving conflicts does not include adding new contacts (because a new contact is not a conflict – it can just be added).

## Auto-Update-Check

The library does contain an update check and will inform you by a log entry if there’s a new version of the library available.

## Planned things

The items listed here are planned, but not promised. Also I don’t have a fixed order in which I will do the planned items.

1. Authenticating proxy support for Xing-Connector … I have a few locations where I cannot use my http-library, because there is a proxy that needs authentication.
2. Authentication and authorization for the web service … I hope to be able to host the web service on my site – some day.
3. Better code … there are some things in the current code that have been implemented to quickly go forward – I will clean up that code and also plan to document all code.
4. More connectors … I will write connectors for Active Directory, Web.de and other social network sites. You have the code, you might be able to implement some by yourself – please send me the code if you write some implementations so that I can use that code in this project and publish it. If you send code to me and I publish it, I will add comments that will clearly state that you did submit the code. But if you send code to me that I already have (implemented by myself or sent by another person) you will not be mentioned. If you don’t like this rule: don’t send me code.

## FAQ

Question: Why does a “sync” update so many contacts in Outlook every time?

Answer 1: I currently do not have a good picture comparison, so if there is a picture in the source, it will always override the one in the target.

Answer 2: Outlook does not accept the picture binary, but converts the specified picture before saving it to its internal storage, so extracted pictures from Outlook are different to the source images that have been imported.