## Общие положения

Для импорта данных реализована подсистема загрузки данных из CSV файла.

Подсистема работает в 2-х режимах:

1. Загрузка данных с помощью remote вызова метода сервиса ImportDataService.
2. Автоматическая загрузка данных при старте сервера.

### Загрузка данных с помощью remote вызова метода сервиса ImportDataService.

Интерфейс сервиса загрузки данных:

**package** ru.intertrust.cm.core.business.api;

/\*\*

\* Сервис загрузки данных

\*

\*/

**public** **interface** ImportDataService {

**public** **static** **final** String *TYPE\_NAME* = "TYPE\_NAME";

**public** **static** **final** String *KEYS* = "KEYS";

**public** **static** **final** String *EMPTY\_STRING\_SYMBOL* = "EMPTY\_STRING\_SYMBOL";

/\*\*

\* Удаленный интерфейс

\* **@author** larin

\*

\*/

**public** **interface** Remote **extends** ImportDataService{

}

/\*\*

\* Метод загрузки данных из файла.

\* **@param** loadFileAsByteArray зачитанный массив данных из файла

\*/

**void** importData(**byte**[] importFileAsByteArray);

/\*\*

\* Метод загрузки данных при отличие кодировки в CSV от дефалтовой ANSI-1251

\* **@param** importFileAsByteArray

\* **@param** encoding

\*/

**void** importData(**byte**[] importFileAsByteArray, String encoding);

/\*\*

\* Метод загрузки данных при отличие кодировки в CSV от дефалтовой ANSI-1251 и флагом который позволяет\запрещает перезаписывать данные

\* **@param** importFileAsByteArray

\* **@param** encoding

\*/

**void** importData(**byte**[] importFileAsByteArray, String encoding, **boolean** owerwrite);

}

Для загрузки данных необходимо вызвать метод importData сервиса ImportDataService и передать методу как параметр массив байт зачитанного CSV файла. Кодировка файлов по умолчанию должна быть windows-1251. В случае иной кодировки ее надо передать в качестве параметра encoding. Параметр owerwrite указывает перезаписывать или нет строку из csv файла если найдена аналогичная по ключевым полям. Если указан owerwrite=true то строка перезаписывается если изменилось любое поле в csv, если owerwrite=false строка не перезаписывается даже если поля в csv отличаются от их значений в базе.

### Автоматическая загрузка данных при старте сервера

При старте сервера производится поиск всех файлов META-INF/cm-module.xml. Файлы могут быть расположены в разных архивах. В файлах cm-module.xml, наряду с другой информацией, содержатся имена файлов csv для загрузки и зависимость от других модулей, для обеспечения корректной очередности загрузки файлов. Зависимость учитывается иерархически. Так же производится проверка на зацикленность зависимостей, в этом случае формируется исключение. Формат фрагмента файла cm-module.xml для указания загружаемых файлов:

<?xml version="1.0" encoding="UTF-8"?>

<module>

<name>test-sed</name>

<description>Тестовый модуль системы электронного документооборота</description>

<depends>

<depend>core</depend>

</depends>

<import-files rewrite="true" csv-encoding=”windows-1251”>

<import-file>importdata/f1.csv</import-file>

<import-file rewrite="false">importdata/d1.csv</import-file>

<import-file>importdata/c1.csv</import-file>

</import-files>

</module>

Тег depends.depend служат для опеределения зочередности загрузки csv файлов, сначала грузятся все модули от которых зависит текущий модуль с учетом иерархии модулей.

В теге import-files указываются общие параметры импорта:

Rewrite – параметр необходимости перезаписывать существующую строку

csv-encoding – кодировка csv файлов, по умолчанию windows-1251.

В теге import-file есть фозможность переопределить параметр rewrite для одного конкретного csv файла в атрибуте rewrite.

### Формат csv файла для импорта данных

По умолчанию файл CSV должен сохранен в кодировке ANSI-1251. Сделано это для совместимости с редактором Excel. В случае если кодировка файла отлична от ANSI-1251 кодировка должна быть передана с помощью параметра encoding метода importData, а в случае автоматической загрузки через cm-module.xml кодировка должна быть указана в атрибуте csv-encoding тега import-files.

Две первые строки файла отведены под метаинформацию. В первой строке задается имя типа для импорта, ключевые поля через запятую и символ, который будет интерпретироваться как символ строки с нулевой длиной в виде:

TYPE\_NAME=Department;KEYS=Name,Number;EMPTY\_STRING\_SYMBOL=$

TYPE\_NAME – имя создаваемого типа

KEYS – поля, перечисленные через запятую, по которым производится поиск записей в базе данных. Если в базе найдена запись с совпадающими полями перечисленными в поле KEYS, то данная запись обновляется, если не найдена то создается новая.

Пустая строка интерпритируется как null.

EMPTY\_STRING\_SYMBOL – указывает какой символ считать пустой строкой, Значение EMPTY\_STRING\_SYMBOL по умолчанию равен символу подчеркивания “\_”.

Во второй строке хранятся имена полей для импорта:

Name;Organization;ParentDepartment

Начиная с третьей строки идут непосредственно данные для импорта

Подразделение 1;"name=""Организация 1""";

Подразделение 2;"name=""Организация 1""";

Подразделение 3;"name=""Организация 1""";

В полях типа Reference необходимо прописывать выражение, для получение значения ссылки на другой доменный объект. Выражения могут быть трех видов:

1. field\_name=”field\_value”. Данное выражение означает, что надо получит конфигурацию поля типа Reference из конфигурации доменного объекта. Далее у поля типа Reference получается тип доменного объекта на который ссылается данное поле, изо всех доменных объектов полученного типа делается запрос с условием field\_name=”field\_value” и из полученного множества берется первый результат.
2. type\_name.field\_name=”field\_value”. Данное выражение означает, что изо всех доменных объектов типа type\_name делается запрос с условием field\_name=”field\_value” и из полученного множества берется первый результат. Данная конструкция используется когда поле Reference может ссылаться на несколько типов доменных объектов.
3. Запрос на внутреннем sql языке. Запрос должен возвращать поле типа Reference. Например: select t.id from department t where t.name='Подразделение 2'.

Поля для импорта даты имеют следующие форматы:

TIMELESSDATE - "dd.MM.yyyy"

DATETIMEWITHTIMEZONE, DATETIME - "dd.MM.yyyy HH:mm:ss" таймзона при импорте используется текущая.

Для импорта Boolean значений используются константы true и false

Все строки содержащие кавычки внутри строки должны быть взяты в кавычки слева и справа а кавычки по середине должны заменятся двойными кавычками (поведение excel). Например name="Организация 1" => “name=""Организация 1”"”.