Entering the Digital Customer Onboarding Era
How Semantic Web Can Help

F. Chevalier & S. Ferré - ISWC 2017 Industry Track - 23 to 25 October - Vienna
Plan

- Company/Product Introduction
- Image based ID Document checking
- How Semantic Web Technologies came to the rescue
- A few figures
- Conclusion
- Credits: IDFraud project
Company and Product Introduction
Who is AriadNEXT?

A dynamic start-up:

- 2010 Company Launch
- 30+ Experts & Engineers
- >3 M€ 2017 forecast

3 expertise sectors:

- Documents Checking & KYC
- Dematerialized Subscription (Point of Sales)
- Electronic Seal Stamping
A typical journey to consumer loan

- On the customer side:
  - Paper copy of **ID document/Passport/Driving license** (are you a real person?)
  - Paper copy of **utility bill** (address proof)
  - **Payslip** (do you earn enough? Do you have some outstanding loans?)
  - **Print** loan contract, **Sign It**
  - Send it by **mail**
A typical journey to consumer loan

- On the bank side:
  - **Process** paper documents (takes time, costs money)
  - **Archive** paper documents (paper holds the legal proof)
Why move to digital customer onboarding?

- Moving away from paper world and pen signature:
  - Provide digital copies of all documents (including ID)
  - Digital signatures
- Real-time processing: get a feedback in matter of minutes instead of days.
- Do everything from your smartphone
- Replace humans screening documents by algorithms (costs savings)
IDCHECK.IO analyzes *in real time* and automatically your identity documents

**Send the document**  
01 The picture of the ID to be verified is sent to our servers to be processed

**Analysis**  
02 Our service automatically controls the documents submitted, in real-time

**Result**  
03 You obtain a precise answer on the ID conformity and all the data extracted from it

**Documents verified:**
- Worldwide ID cards, passports, visas, residence permits…
- French documents also verified:
  - Proof of address (utility bill)
  - IBAN
  - Vehicle registration certificate
  - Tax sheet
  - Pay slip

www.ariadnext.com – contact@ariadnext.com
Image based ID document checking
Ultra fast intro by example
Image based ID document checking

Let’s look a ID document for a moment...
Image based ID document checking

You have to tell the computer:

- Where is the text that is to be read
- Relative positioning (this text is on right of this one...)
- Embedding (This date is encoded inside this number)
- What it is expected to read (Charset, Numbers, names, font used..., so that we fine tune the OCR)
- What kind of security features are in place (what can/should be checked?). Where are located those areas?
Image based ID document checking

An example of ID document design complexity: Printing layers.

Document moves from one printer to another… and text starts “floating”
Image based ID document checking

Printing layers example
Other examples of ID design complexity

- Dates (01/12/2017, 12/01/2017, 12 JAN 2017, xx/xx/2017, Japanese)
- Names (First, Last, or both, father, mother…)
- Addresses
- … and much more

=> We discover new complexities every week
How Semantic Web Technologies came to the rescue

Aka help from OWL/RDFS, Protégé, AutoRDF, ModelEditor
Knowledge Management Module Requirements

- **Support high rate data model evolution.** Data model gets refined every week as we discover new innovative ways ID documents designers decide to do the same thing in a different way.
- **Data consistency:** undefined behaviours prevention, as algorithms will be fed with data from our KM module.
- **Easy mapping to OO paradigm:** Need a way to expose concepts in a developer friendly way.
- **Easy data edition:** Should be able to enter data with zero knowledge on database technology used.
- **Long term support:** Best if model and data stored in a standard compliant way, as they will have to be maintained for decades.
Knowledge Management Workflow

New document support request

Ontology design

Reflect changes in code

GUI for document description
Knowledge Management Workflow

New document support request

Ontology design

Protégé

Reflect changes in code

AutoRDF

GUI for document description

ModelEditor
AutoRDF

OWL2 or RDFS

```cpp
class IPoInt { 
public:
    /**
     * @return the mandatory value for this property.
     * @throw PropertyNotFound if value is not set in database
     */
    double x() const {
        return object().getPropertyValue("http://example.org/geometry#x").get<autordf::cvt::
    }

    /**
     * @param value value to set for this property, removing all other values
     */
    IPoInt& setX(const double& value) {
        object().setPropertyValue("http://example.org/geometry#x", autordf::PropertyValue().
        return *this;
    }
}

https://github.com/ariadnext/AutoRDF
A few figures
Database: triples count growth
Business: Turnaround growth

Turnaround in €

- 2014: 0
- 2015: 150,000
- 2016: 750,000
- 2017: 1,500,000
Conclusion

- Our Knowledge Management Workflow saves us a lot of time.
- We spend time designing the best possible ID Document verification schema instead of writing boilerplate Database and GUI code.
- All that would not have been possible without the help of Semantic Web.
Credits: IDFraud Project

- Funded by

- Coordinator: Ahmad Montaser Awal
- Web site (idfraud.fr)
- Partenariat
  - Private company: AriadNEXT
  - Research Lab: UMR IRISA
  - Documents Fraud Experts: Centre de recherche, Bureau de Fraude Documentaire, Département Documents
Extra Slides
IDFraud Project

- An automated ID documents analysis framework
  - Worldwide coverage
  - Extensible
  - Documents checking, security patterns
  - Works with scanners, smartphones, … or whatever makes images from ID documents
  - *Fraud profiling*
IDFraud Technical Overview

ID Document Analysis

ID Document Knowledge Management

ID Fraud Pattern Extraction

Public registers

Fraud patterns

Document dataset
Image based ID document checking

Checksums and fonts
Image based ID document checking

Initials

FH
Image based ID document checking

MRZ Structure & Font
Classification & Crop