Technology
Arts Sciences
TH Köln

Improving E-commerce fraud investigations in virtual, inter-institutional teams:

Towards an approach based on Semantic Web technologies

Master thesis

by

Andreas Gerlach

submitted to obtain the degree of

Master of Science (M.Sc.)

at

TH Köln - University of Applied Sciences
Institute of Informatics

Course of Studies
Web Science

First supervisor: Prof. Dr. Kristian Fischer
TH Köln - University of Applied Sciences

Second supervisor: Stephan Pavlovic
TH Köln - University of Applied Sciences

Cologne, August 2016

Contact details:

Andreas Gerlach
Wilhelmstr. 78
52070 Aachen
andreas.gerlach@smail.th-koeln.de

Prof. Dr. Kristian Fischer
TH Köln - University of Applied Sciences
Institute of Informatics
Steinmüllerallee 1
51643 Gummersbach
kristian.fischer@th-koeln.de

Stephan Pavlovic
TH Köln - University of Applied Sciences
Institute of Informatics
Steinmüllerallee 1
51643 Gummersbach
stephan@railslove.com

Abstract

There is a dramatic shift in credit card fraud from the offline to the online world. Large online retailers have tried to establish countermeasures and transaction data analysis technologies to lower the rate of fraudulent transactions to a manageable amount. But as retailers will always have to make a trade-off between the *performance* of the transaction processing, the *usability* of the web shop and the overall *security* of it, we can assume that E-commerce fraud will still happen in the future and that retailers have to collaborate with relative parties on the incident to find a common ground on and take coordinated (legal) actions against it.

Combining information from different stakeholders will face issues due to different wordings and data formats of the information, competing incentives of the stakeholders to participate on information sharing as well as possible sharing restrictions, that prevent making the information available to a larger audience. Additionally, as some of the information might be confidential or business-critical to one of the involved parties a *centralized* system (e.g. a service in the cloud) could **not** be used.

This Master thesis is therefore looking into the topic of how far a computer supported collaborative work system based on peer-to-peer communication technologies and shared ontologies can improve the efficiency and effectivity of E-commerce fraud investigations within an inter-institutional team.

Keywords: Peer-To-Peer Communication, Semantic Web, CSCW

Contents

1 Introduction

This introductory section of the Master thesis will first give a section showing the importance and relevance of the topic in the research area of Web Science, followed by a description of the problem this thesis will focus on as well as an analysis of related works and an overview of the outline of the thesis.

1.1 Motivation

> *"When it comes to fraud, 2015 is likely among the riskiest season retailers have ever seen, [...] it is critical that they prepare for a significant uptick in fraud, particularly within e-commerce channels."*

This statement from Mike Braatz, senior vice president of Payment Risk Management, ACI Worldwide in (Reuters 2015) shows the dramatic shift in credit card fraud from the offline to the online world, that retailers are starting to face nowadays.

In general credit card fraud can occur if a consumer has lost her credit card or if the credit card has been stolen by a criminal. This usually results in an **identity theft** by the criminal, who is using the original credit card to make financial transactions by pretending to be the owner of the card. Additionally, a consumer might hand over her credit card information to an untrustworthy individual, who might use this information for her own benefit. In the real world scenario there is usually a face-to-face interaction between both parties. The consumer, wanting to do business with a merchant or interacting with an employee of a larger business, has to hand over her credit card information explicitly and can deny doing so if she faces a suspicious situation. The criminal on the other hand must get access to the physical credit card first, before she is able to make an illegal copy of it — a process called **skimming**. The devices used to read out and duplicate the credit card information are therefore called skimmers. These can be special terminals, that the criminal uses to make copies of credit cards she gets her hands on, or they can be installed in or attached to terminals the consumer interacts with on her own (Consumer Action 2009). All of these so-called *card-present transaction* scenarios have seen a lot of improvements in security over the last years. Especially the transition from magnetic swipe readers to EMV chip-based credit cards makes it more difficult for criminals to counterfeit them (Lewis 2015).

As of this criminals are turning away from these card-present transaction scenarios in the offline world. Instead they are focusing on transactions in the online and mobile world, in which it is easy to pretend to own a certain credit card. Most online transactions (either E-commerce or M-commerce) rely ***only*** on credit card information like card number, card holder and security code for the card validation process – as of this these interactions are usually called *card-not-present transactions*. This credit card information can be obtained by a criminal in a number of ways. First she might send out **phishing emails** to consumers. These emails mimic the look-and-feel of emails from a merchant or bank, that the consumers are normally interacting with, but instead navigating the consumers to a malicious web site with the intend to capture credit card or other personal information (Consumer Action 2009). Additionally, criminals can **break into the web sites** of large Internet businesses with the goal of getting access to the underlying database of customer information, that in most cases also hold credit card data (Holmes 2015). Additionally, some of the online retailers are not encrypting the transaction information before transmitting them over the Internet; a hacker can easily start a **man-in-the-middle attack** to trace these data packages and get access to credit card and/or personal information in this way (Captain 2015).

Based on this it should come not as a surprise that the growth rate of online fraud has been 163% in 2015 alone (PYMNTS 2016). This results in huge losses for the global economy every year and it is expected that retailers are losing $3.08 for every dollar in fraud incurred in 2014 (incl. the costs for handling fraudulent transactions) (Rampton 2015). These fraudulent transactions also impact the revenue of the online retailers. Here we have seen a growth of 94% in revenue lost in 2015. Overall it is estimated that credit card fault results in $16 billion losses globally in 2014 (PYMNTS 2016) (Business Wire 2015).

While it is possible to prevent fraudulent transactions in the card-present real-world scenario (mainly due to introducing better technology and establishing organizational countermeasures in the recent past), it is more difficult to do so in the card-not-present online- and mobile commerce scenarios, which are lacking face-to-face interactions and enable massive scalability of misusing credit card information in even shorter time frames (Lewis 2015). Large online retailers have tried to establish countermeasures and transaction data analysis technologies to lower the rate of fraudulent transactions to a manageable amount. But this is still an expensive and inefficient solution to integrate into the retailers' business processes, and is largely driven by machine-learning techniques and manual review processes (Brachmann 2015). Additionally, it can be assumed, that the online retailers are getting into a Red Queen race with the criminals here: with every new technology or method introduced they might just be able to safe the status quo. This is largely due to the facts, that there will be no 100% security for such a complex and interconnected system like an E-commerce or M-commerce shop, the criminals will also increase their efforts and technology skills to adapt to new security features and most importantly retailers will always have to make a trade-off between the *performance* of the transaction processing, the *usability* of the web shop and the overall *security* of it.

1.2 Problem Definition

This Master thesis will **not** look into novel techniques and methods to *prevent* credit card fraud in the E-commerce world. This aspect has been seeing a lot of research in the last years.[1] Instead this Master thesis will look into a **concept to optimize the collaboration** between the affected stakeholders in case of an existing credit card fraud in an E-commerce system.

Stakeholders might include **vendors** and other businesses, that the retailer has a long-term business relationship with, **law enforcement agencies**, **payment service providers** like PayPal or Visa, **banks**, and even **competitors**, that are also affected by the Internet fraud. In such a case the merchant usually tries to solve the issue on his own and getting in contact with relative parties by phone or e-mail if necessary. But these communication styles do not fit to the complexity of the task involved, and based on the media-richness model (see Figure 1.1) will result in inefficient and ineffective problem solutions.

Due to the task complexity a **physical face-to-face meeting** with representatives of all involved stakeholders might be a good fit, but arranging such a meeting (same time, same place) with multiple parties, that are globally dispersed, is either economically not feasible or takes a lot of time. But the more time passes for investigating the crime the more difficult it will become to find the criminals and take legal actions against them, which can also reduce the risk of losing the stolen money completely.

[1] please also note the various US patent applications of Google on that matter from 2015, e.g.: "Credit card fraud prevention system and method", "Financial card fraud alert", "Payment card fraud prevention system and method" (Google Patents)

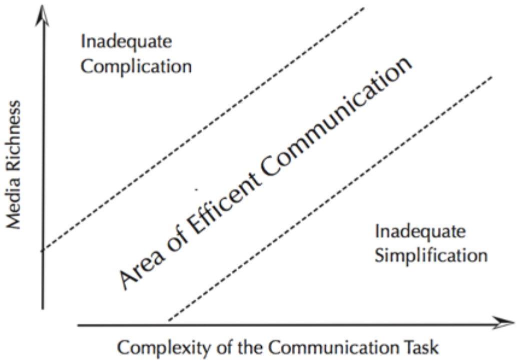

Inadequate Complication
Media Richness
Area of Efficent Communication
Inadequate Simplification
Complexity of the Communication Task

Figure 1.1: The Media Richness Model (Rice 1992)

As of these conditions a **computer-supported collaborative work** (CSCW) system might be an alternative to *cooperate* on an incident of E-commerce fraud (same time, different place). CSCW systems can be categorized by their support for the mode of group interaction as done in the 3C model:

- **communication:** two-way exchange of information between different parties
- **coordination:** management of shared resources like meeting rooms
- **collaboration:** members of a group work together in a shared environment to reach a goal

Based on the level of support for one of these functionalities the various systems can be classified and described (see Figure 1.2) (Koch 2008):

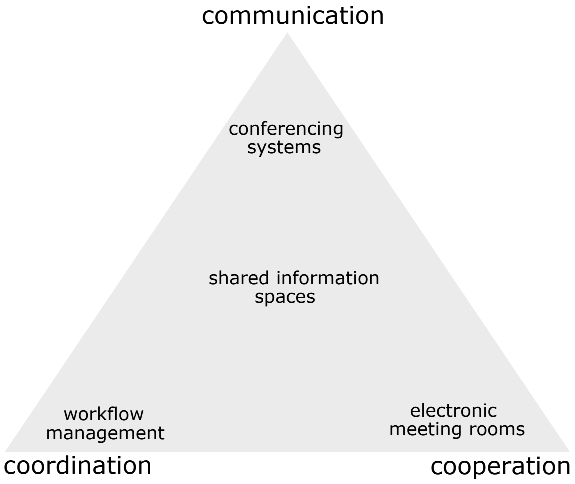


Figure 1.2: The 3C Model (Koch 2008)

A good candidate *could* be a **shared information space**; aka team rooms, cloud storage services or document management systems, that allow to access information at any place, any time and to share information with co-workers — usually with a build in versioning support for artefacts and a workflow component.

However as some of the required information might be confidential or business-critical to one of the involved parties a **centralized system** (e.g. a service in the cloud) could **not** be used in the scenario described here. Another key characteristic of the investigation of an E-commerce fraud is, that it involves information sharing from many different organizations. These different aspects have to be combined into a **shared information space** in a meaningful way to be able to achieve the common group goal on time. Combining information from different stakeholders will face issues due to **different wordings and data formats** of the information, **competing incentives** of the stakeholders to participate on information sharing as well as possible **sharing restrictions**, that prevent making the information available to a larger audience.

Decentralized information sharing architectures, that utilizes **peer-to-peer communication technologies**, are either restricted to a commonly agreed set of data entities and relations (based on an ontology) between all involved parties or are lacking richer semantics for sharing and integrating content between the stakeholders. **Semantic Web technologies** can help lower the barrier to integrate information from various sources into a shared information space, and the advantages of peer-to-peer communication and Semantic Web technologies for information sharing in distributed, inter-organizational settings have been shown in (Staab & Stuckenschmidt 2006).

Still these studies concentrate on making information from different parties searchable and accessible in a distributed, shared information space, which data can be accessed and queried at any time from any participating party. They are not solving the problem of working collaboratively on a common goal in an ad-hoc, loosely-coupled virtual team of disperse organizations by making certain (sometimes sensitive) information available in a shared environment.

Therefore, the **research question** for this Master thesis can be summarized as:

> *In how far can a computer supported collaborative work system based on peer-to-peer communication technologies and shared ontologies improve the efficiency and effectivity of E-commerce fraud investigations within an inter-institutional team?*

1.3 Master Thesis Outline

In the first section this Master thesis will discuss the E-commerce scenario in detail. It starts with a description of the E-commerce shopping process, looks into the stakeholders involved and further shows possible kinds of E-commerce fraud. Based on this discussion the chapter will close with a presentation of the specific scenario selected for the further investigation within this Master thesis.

After this initial scope setup the thesis will briefly outline the theoretical foundations required for the understanding of the concepts in the solution space. This section starts with a short overview of the importants aspects of computer-supported collaborative work systems, shows the needed technical specifications of the Semantic Web standards and ends up with an introduction to the peer-to-peer communication techniques and protocols.

Before starting with the investigation of possible solutions for the problem described in the end of the first section the thesis will also list related works, that has been examined in the course of this Master thesis and have had an influence on the solution space.

Last but not least the conceptualization of a collaborative system that supports the investigation of E-commerce fraud takes place. This chapter will lay out and discuss the possibilities for designing and using such a system. The objective is to come up with an approach at the end of this chapter, that might be the best fit for the problem described in the first section.

To conclude the thesis will also show an outcome of the paper work as well as give an outlook that might be useful to decide future progress on this topic.

2 Context Analysis

This chapter will look into the scenario of E-commerce fraud investigation in detail. It will start with an in-depth scenario description followed by an analysis of the stakeholders involved. It will further describe the kind of information each stakeholder has in her local context and her objectives to take part on the information sharing and collaboration initiative. Based on the analysis of the possible kinds of E-commerce fraud, the chapter will end with a description of the selected scenario for this Master thesis.

2.1 Scenario Description

E-commerce as a term relates to the trading of products or services utilizing a computer network such as the Internet. It is usually categorized into the following four different subfields (Sen et al. 2015):

1. **Business-To-Business (B2B)**: refers to electronic trading between companies with the objective to improve their supply chain processes
2. **Business-To-Consumer (B2C)**: refers to electronic trading between a company and it's consumers (most publicly known example is Amazon)
3. **Consumer-To-Consumer (C2C)**: refers to electronic trading between consumers (most publicly known example is eBay)
4. **Consumer-To-Business (C2B)**: referes to electronic trading between consumers and businesses (most publicly known example is TaskRabbit)

This Master thesis will solely focus on the **B2C** aspect of E-commerce. In that case a **consumer** is using an **E-commerce shop** of a **merchant** on the Internet to order products or services online. The merchant is offering a catalog of available products or services on the Web, that is available and accessible by the general public and usually has an at least nation-wide if not global reach. The merchant can either run the E-commerce shop software on her own servers (on-premise) or can outsource this additional sales channel to a **3rd party hosting company or cloud service provider**

(CSP). Also the E-commerce shop software itself can either be developed by the merchant internally or acquired as a boxed product on the market from an **Independent Software Vendor (ISV)**. For business accounting purposes the merchant also runs a bank account with the **acquirer** (see Figure 2.1).

When placing an order with the merchant online the consumer is normally using a **credit card** for finalizing the transaction. This credit card has originally been handed out by the **issuing bank** to the consumer. Additionally some online shops make it mandatory to the consumer to create an user account with them while others do not. The former is the preferred way when consumers are repetitively buying from that merchant whereas the latter might be used for one-time or irregular shopping trips online. Last but not least to be able to connect to the Internet the consumer relies on a service of an **Internet Service Provider (ISP)**. The whole initial setup for participating on E-commerce activities is found in Figure 2.1.

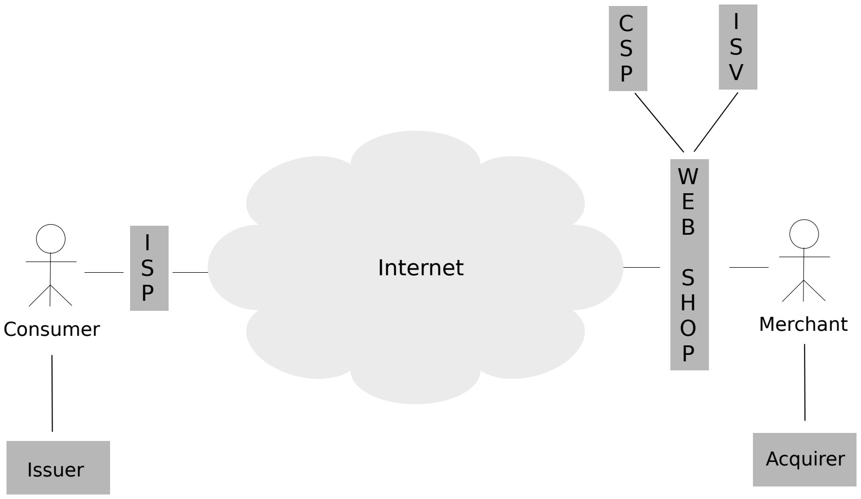


Figure 2.1: E-commerce Fundamentals

When the consumer places the **order** online, the merchant receives at least a **list of products or services** from the current shopping cart of the consumer as well as the **delivery address** to ship the physical items to. If the transaction is going to be finalized with a credit card, the consumer will have to state additional information like her **billing address** and **credit card information** (including the credit card number, the expiration date and a security number).

The merchant does not validate the credit card information on her own. For that purpose she is usually relying on another 3^{rd} party service offered on the Internet by the **Payment Service Provider (PSP)**. These providers are either validating the credit card information themselves based on an user profile the consumer has with the PSP (e.g. a central Web service like PayPal) or connecting to the issuing bank of the card for doing so. For this validation process the merchant is handing over the following information to the PSP:

- consumer's billing address
- given credit card number, expiration date and security number
- identification of the merchant
- final amount of the transaction actually being processed

Either the PSP or the issuing bank is validating the correctness of the information with criterias like:

- is the billing address matching the current consumers' postal address on file?
- is the stated credit card information correct and is this credit card not marked as being blocked?

The merchant will receive the **status** of the authorization as well as a **payment token** in return. If the authorization was done successfully the merchant will collect the items and send out a shipping request to one of the available **logistic services** capable to handle the order. They will pickup the order at the merchant's facility and ship it to the delivery address stated by the consumer. Usually in parallel the merchant is informing her bank about the order, amount due as well as payment token from the PSP. The acquirer is in charge to withdrawal the amount of the order from the consumers bank account either via the PSP or directly with the issuing bank depending on who of them has authorized the initial payment request (a process called clearing) (Visa Europe 2014).

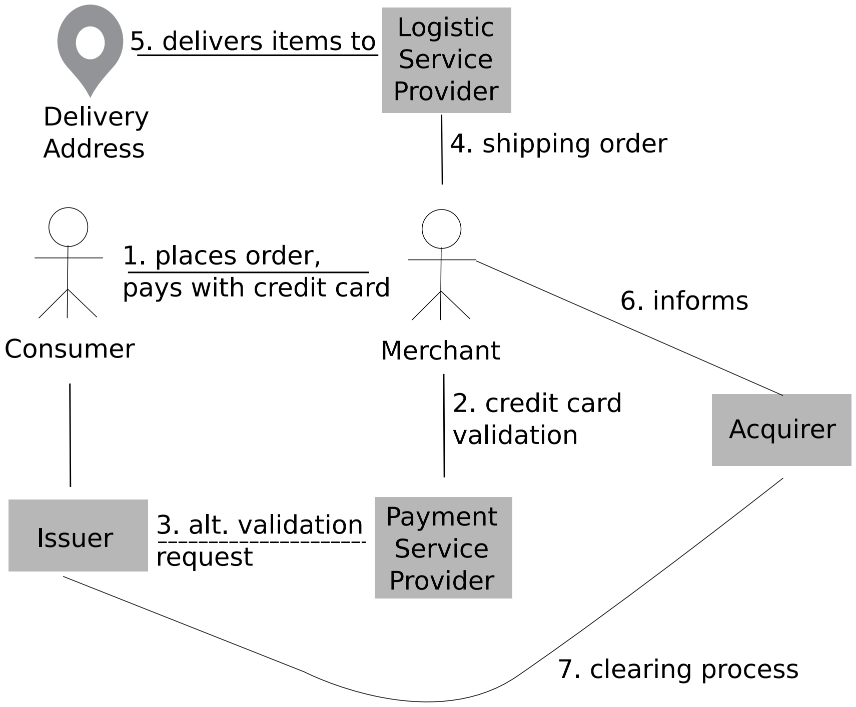


Figure 2.2: E-commerce Checkout Process in detail

2.2 Stakeholder Analysis

The following section will look at each stakeholder involved in detail, listing the kind of information they have at hand and the incentives they might have to participate on a shared collaborative system.

2.2.1 Consumer

The **consumer** is the initiator of an E-commerce transaction. She is using the shop of a **merchant** on the Internet to order products or services. For this she has to know the URL of the Web shop, has to be connected to the Internet via the **ISP** and have to use a standard software called a Web Browser on her computer or mobile. For the duration of her online session she receives an unique **IP** address from the ISP. She might have had a long-term business relationship with the merchant and due to this owns an **user account** on the Web shop. On the other hand she might be just interested into a one-time shopping trip and might want to order the items without creating an account first — something called "anonymous shopping" in the E-commerce scenario.

The consumer is also having a **bank account** and at least owns a **debit card** with the **issuing bank** to get access to the money on that account. In addition to that she can also hold multipe **credit cards**. A credit card can be issued by the same bank or can be provided by other financial services (e.g. American Express). In any case the organisation that has handed out the credit card to the consumer is called the **issuer**.

If she is going to order items in a Web shop she will usually browse the product and service offerings of the merchant first and put the articles of interest into the **shopping cart**. When finalizing the transaction she has to hand over the following information to the merchant:

- personal information incl. given name, family name and date of birth
- the current home address
- the address the items should be shipped to
- payment information incl. type of payment and billing address

If she is going to end the transaction with a payment of type **credit card** she will have to provide the specific information of the card to be used:

- credit card number
- credit card expiry date (in format MM/YY)
- credit card security code

The consumer is also playing a special role in the whole scenario. As the merchant has to deal with the consumer without any face-to-face or real-world interaction, the consumer is also the less thrustworthy party from the point of view of the merchant. As the Section 2.3 will show the consumer is the main object of question in the case of an E-commer fraud. For the investigation of it the consumer is usually not taking an active part.

2.2.2 Merchant

The **merchant** offers products and services on the Internet for the general public. She might use the Internet as an additional sales channel or rely on it solely for making any business. To provide access to the **Web shop** the merchant has to register a domain name and **URL** at a local registry. This specific URL refers to a fixed public

IP address, that the server hosting the Web shop software uses. Normally the merchant does not run the servers herself, but rely on a service offering from a hosting or **cloud service provider** for that. Also the Web shop software itself is not implemented by the merchant, but bought from an **ISV**. The merchant has special rights in this software as she is allowed to configure the products, prices, promotions, available payment and shipment services. Products can be categorized into departments and sub-departments for easier navigation in the online shop.

The merchant can decide whether she restricts ordering of products to registered users or allows anonymous users too. The main benefit of the former is the possibility to analyse the **shopping behaviour** of individual consumers, whereas the latter will open the business for a wider range of consumers as it include also those, that do not want to register with any online shop. Nevertheless any consumer activity on the online shop is traced in the analytic databases of the merchant. This includes not only the items, that have been placed into the shopping cart, but also any product that has been looked at during this shopping session. Even if these detailed analytic capabilities are actually synonymous for target-related advertising, they can also help to decide if a consumer behaves normally or not.

Any business transaction that the consumer makes with the merchant is stored in the merchants' databases. A transaction information contains, but is not limited to:

- the personal information of the consumer
- the address the items will be shipped to
- a collection of products that have been ordered
- the total amount of the order considering promotions, taxes and fees
- the selected payment information

If the consumer pays with credit card the merchant do not handle the payment herself, but relate this action to a **payment service provider**. In return of the payment authorization the merchant will receive and store the following payment-related information for the transaction:

- the type of credit card uses (e.g. Visa, MasterCard, American Express, ...)
- the credit card information incl. credit card number, expiry date and security code

- the name of the credit card owner
- the payment token received by the payment service provider
- the timestamp and result code of the authorization
- the authority who approved the payment (if the merchant offers multiple payment service providers)

As a merchant will collect a lot of personal and payment related information over time, she is also one of the major sources of possible data leaks in this scenario. Due to this fact the Payment Card Initiative provides rules and guidelines (aka PCI/DSS standards) for securely handling these kinds of information in an IT system (DSS 2014).

A merchant is one of the main actors in the fraud investigation process. She is interested in figuring out whether the consumer's transaction is valid or not. As in a case of an E-commerce fraud the merchant will mostly have to cover the costs of the incident (see Section 2.3). Also the online merchant's reputation will suffer if private information from her databases get leacked. If a merchant falls victim to a fraud incident multiple times the economic damages can finally result in a bankruptcy of the merchant.

2.2.3 Payment Service Provider

The **Payment Service Provider** is offering payment-related services to online merchants. As of this the PSP provides a common Web interface to communicate with for payment authorization requests. The PSP might be able to authorize the payment request on her own or have to route the request to the corresponding **issuer** of the credit card in question. For the former procedure to take place the PSP usually has an own database of registered users with their credit card information (e.g. PayPal). For checking the credit card and authorizing the payment the **merchant** is sending the following information from the transaction:

- credit card owner name
- credit card number
- credit card expiry date
- credit card security number
- identification of the merchant

- current transaction total amount

The PSP has to securely process these information and return the result of the examination to the merchant. The result message also contains an unique payment token, that the merchant can refer to later to initiate the clearing process. As of this the PSP have to persist the credit card and transaction related information in his backend databases. Following industry standards she should do so according to the PCI/DSS guidelines mentioned above.

The level of activity in the E-commerce fraud investigation process depends on whether the PSP authorizes the payment herself or only acts as routing service between the merchant and the original credit card issuer. In the former case the PSP is more actively involved. In that case she also holds more of the valuable information to solve the incident. In the latter case she might be able to connect the payment-related request information from the merchant with the authorization result coming from the issuer.

If the PSP holds sensitive information in her own databases she might also be a source of a possible data leak.

2.2.4 Issuing Bank

The **issuer** is one of the parties in the scenario that knows the owner of the credit card in person. Each individual has to register personally with the issuer to get access to a credit card. This includes providing the following information:

- personal information like given name, family name and date of birth
- the current home address
- the bank account that should be used to settle credit card balances

Even if the two parties do not really meet each other personally, a person will still have to identify with a valid id card and bank account to receive and activate a new credit card. Beside being the single source of truth about the original credit card owner the issuer of the card also collects and stores all usages of it. The issuer therefore can provide credit card usage patterns, that are not just limited to the online shopping scenario — something a Payment Service Provider might also be able to deliver; but also include transactions the card owner does in the real-world. Needless to say that these are valuable information for the E-commerce fraud investigation.

Still the issuer do not know the details of each transaction that has been made with the credit card. As mentioned above in the section about the PSP the issuer will just receive an identifier of the merchant, in which shop the credit card has been used. But based on public available information of the merchant from the commercial register the issuer could at least come up with the retail branch the merchant belongs to.

Being the single source of truth about all issued credit cards and their owners the issuer is another high-risk for data leaks. They should as well follow the guidelines from the PCI/DSS standards and monitor their backend systems heavily for intrusion detection.

2.2.5 Acquiring Bank

The **acquirer** holds the bank account of the merchant and is responsible for withdrawing the outstanding amounts of transactions from the accounts of the consumers, or more precisely the issuing bank of each consumer. As of this the acquiring bank is usually not processing any credit card related information from consumers, but refers to the payment tokens that have been given by the PSP or issuer during the authorization process.

Still as financial institute it has to comply with the rules and guidelines of the PCI/DSS and other industry standards to make sure that their bank accounts and the transaction processing are safe and secure. The detailed analysis of these procedures and their possible (banking) frauds is out of scope of this thesis.

2.2.6 Logistic Service Provider

The **Logistic Service Provider** has two important roles in the E-commerce scenario. First it has access to and control over the items of the merchant for the duration of the transport between the merchant's facility and the consumer's shipping address. And second it holds the information to whom it has handed over the items at the final destination. Although the LSP has nothing to do with any payment related activities, it might be the last chance for the merchant to stop the delivery of the items (in case a fraud has been detected after shipment) or give information about the person that has received the items at the shipping address — especially on high-priced goods, which usually require the receiver to show their personal id card and have to place a signature on the delivery receipt.

For initiating the shipment procedure the merchant is ordering a certain transport service from the LSP and hand over the following information:

- name of the receiver
- delivery address
- list of items to be shipped
- optionally: value of the items if an insurance policy is taken

The LSP at the other hand returns an unique **tracking id** for the shipment. It can be used by the merchant and the consumer to check the status of the shipment online.

2.2.7 Cloud Service Provider

The **Cloud Service Provider** offers IT services to its customers. These IT services include hardware and software assets that in our scenario a merchant can order to run her Web shop on the Internet. Part of the service level agreement between the merchant and the CSP is a detailed listing of the responsibilities of both parties (who has to take care of what). In most cases the merchant is outsourcing the complete operation of the hardware and software for the Web shop to the CSP, making the CSP be responsible for making sure the Web shop is available and secure. The CSP is also constantly monitoring the incoming connections to the public Internet servers under her control and might provide information if the Web shop of one of the merchant using the services of the CSP has been compromised or leaked.

2.2.8 Independent Software Vendor

The **ISV** designs, implements and sells the Web shop software. It has detailed knowledge about the software components and libraries used within the Web shop and normally also monitors for security breaches or vulnerabilities in them. It also has to check the own implemented software code for these and additionally have to make sure that the implementation follows industry standards (e.g. PCI/DSS for handling person and payment related information). As of this it can best assert these quality criterias of the Web shop if needed.

2.2.9 Internet Service Provider

The **ISP** provides a service to the consumer for connecting to the Internet. Each Web request the consumer is doing on her system is routed to the public Internet via the infrastructure of the ISP. Due to regulations and laws the ISP have to store the log files of any Internet session of its customers for a certain amount of time. Especially these log files can be helpful to decide whether a consumer was visiting pages in the dark-side of the Web or if she falls victim to some phishing attacks (explained in detail in Section 2.3).

2.3 E-commerce fraud scenarios

Workshop ErsteBank Wien:
- usually banks are monitoring the usage of credit / debit cards and are looking for suspicious activities
- this fraud prevention mechanism is working on a rule-based (non self-learning) or score-based (self-learning) software from a third party
- the outcome of the fraud prevention could be: yes (this is a fraudulent transaction, pls. block it), no (everything looks fine, pls. continue with the process) or maybe (uncertain, pls. let a human decide how to proceed)
- in case of the maybe result an alert is triggered to one of the support staff of the bank (operating 24/7/365)
- still this kind of fraud prevention mgmt. can not solve all issues due to the amount and frequency of the transactions, there is generally a fraud-to-sales ratio of max. 0.11 percent in the EU (meaning 1 promille of transactions are fraudulent)
- still the success rate of fraud prevention is rougly 70-80 percent, means of the fraudulent transactions nearly 80 percent are blocked correctly
- for the rest: the consumer has to actively trigger an investigation, if the case is valid usually the issuing bank will cover the cost (in case of larger amounts an insurance will take over).
- the bank is responsible for pushing the consumer to file a case at police
- most of the filed cases could not be resolved
- there is no court decision yet in which circumstances the consumer might be guilty as well

- ca. 85 percent of frauds are E-commerce frauds (EU: 70-90 percent). Hotspots are Germany, France and US. Frauds are coming from Travel-Shops or Online Merchants and the amount is on average between 500-600 EUR
- E-commerce frauds will usually not filed at police; in most cases the acquirer is in charge to handle the issue
- if it is known that a merchant has been hacked the bank is usually issuing new credit cards to all affected consumers automatically
- otherwise the bank has to get in contact with the acquirer and/or merchant to figure

out if the transaction is fraudulent
- usually the banks only have credit card related information from a consumer (no detailed information about the ongoing transaction), whereas the merchant and the acquirer have the detailed records of the order at hand
- various regulations make it hard to shre detailed information with involved parties (even if they have special agreements signed between them)
- main questions for E-commerce fraud: who is the party that is the victim of the incident? Is it really a fraudulent transaction?
- E-commerce fraud can not be handled by technology alone, at best the fraudulent transaction can be blocked on the merchant side (due to the information given by the consumer like items, prices, delivery address, ...)
- in the worst case one successful fraudulent transaction in an E-commerce shop will trigger hundred and thousands of attempts -¿ so the awareness for the issue has to be at merchant side
- at the end: much effort is assumed to bring all the experts together and solve the issue by putting their individual know-how on the table

2.4 Scope of this Master Thesis

3 Theoretical Foundations

ca. 25 pages

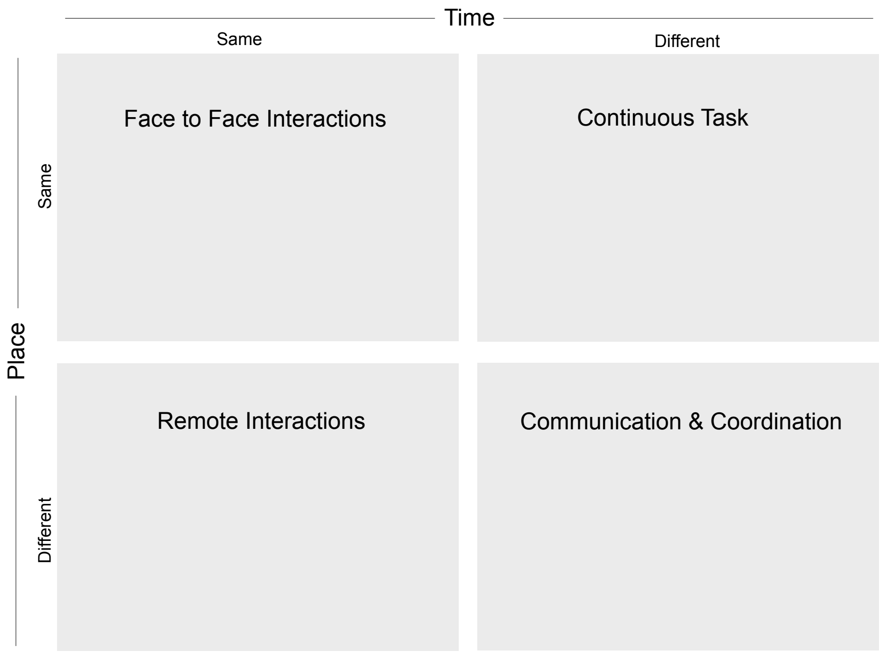
This chapter will lay out the theoretical foundations for the to-be-designed collaborative system. It will start with an investigation of the CSCW system theory followed by a detailed examination of the Semantic Web standards like RDF, OWL and SPARQL and how they can be used within Semantic Web agents. Last but not least the chapter will look into the concepts of P2P communication technologies by looking into various protocols for information sharing in detail — e.g. XMPP, WebRTC as well as less known ones like BitTorrent and BitMessage.

3.1 Computer-Supported Cooperative Work

3.1.1 Definition

3.1.2 Types

CSCW systems can be differentiated by their support of communication on the two axis place and time:



		Time	
		Same	Different
Place	Same	Face to Face Interactions	Continuous Task
	Different	Remote Interactions	Communication & Coordination

Figure 3.1: CSCW Place/Time Matrix (**?**)

Additionally it is possible to group the CSCW systems based on the 3C model:

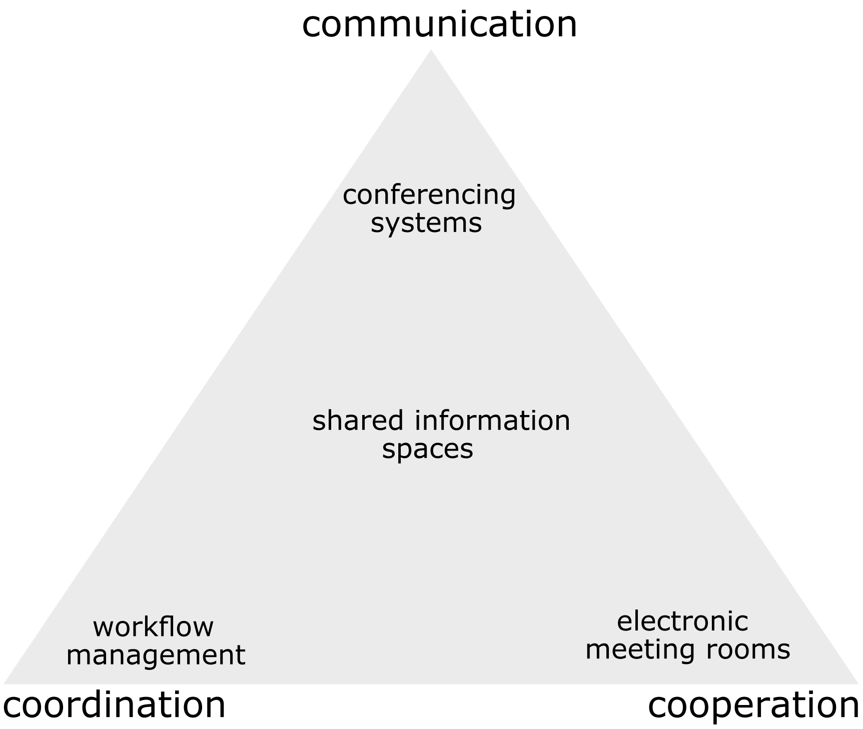


Figure 3.2: The 3C Model (Koch 2008)

3.1.3 Shared Information Spaces

3.1.4 Important aspects of CSCW systems

3.2 The Semantic Web

3.2.1 Vision

MKP Chapter 1:
integrate distributed data from various publishers on the Web into smart applications
the Semantic Web delivers the infrastructure for this vision in form of various standard specifications (RDF, RDFS, OWL, SPARQL, ...)
the fundamentals of the World-Wide Web are also supported by the Semantic Web, especially:
- AAA-Slogan: Anyone can say Anything about Any topic
- Open World Assumption: we must always assume that there exist new information unknown to us yet, that can give additional insights
- Non-unique Naming Assumption: different URIs might refer to the same entity or object

as of this any one can extend on existing data entities and contribute her own knowledge / opinions as well as combine existing information in new ways -¿ data wilderness, no common data schema, more of an organic, living system
it heavily depends on the "network effect" and will / might explode with rising number of users / applications
as there will be disagreements on all sorts of topics there is no single ontology for the whole Web, but rather multiple ontologies tat can be integrated and utilised

MIT Chapter 1:
make information on the Web accessible to machines
- allows integration of information across web sites
- is also known as the "Web of Data"

design principles:
1. make structured and semi-structured data available in standardized formats
2. make individual data elements and their relationships accessible on the Web
3. describe the intended semantics of the data in a machine readable format

HTML is just for human consumption and a lot of the structures and semantics of the underlying databases is lost in the transformation process
- use labeled graphs as data model for objects and their relationships (objects == nodes, edges == relationships between them)
- formalize the syntax of the graph in RDF (Resource Description Framework)
- use URIs to identify individual data items and relations
- use ontologies to represent semantics of the data items (either lightweight RDF schema definitions or Web Ontology Language are used for that)

RDFS and OWL are meta-description languages allowing to define new domain-specific knowledge representations
they rely on the basic principles of the Web: supporting distributed, decentralized architectures

some new initiatives for standardizing semantics: schema.org and linkeddata.org
initially it was tried to solve the integration issues with XML, but as it is syntactically more machine- readable it lacks the semantic of the data
- as of this RDF is the basic language of the Semantic Web and describes meta-data as well as content

an ontology formally describe a domain based on terms and their relationships (terms == classes of objects)
hierarchies are supported (even multiple inheritance between objects)
ontologies also include:
- properties
- value restrictions
- disjointness statements
- specifications of logical relationships
goal is to provide a shared understanding of a domain
can help with the necessity to overcome differences in terminology
a mapping for different wordings in an ontology or between ontologies is possible
they can also be useful for generalization or specialization of Web search results

ontologies help with reasoning of objects, they can uncover unexpected relationships and inconsistencies as well as - by utilizing intelligent web agents - make decisions and select course of actions (e.g. “if-then-conclusions” aka Horn logic)
agents can also be used for “validation of proof” of statements of another agent or machine

Semantic Web is a layered approach . . .

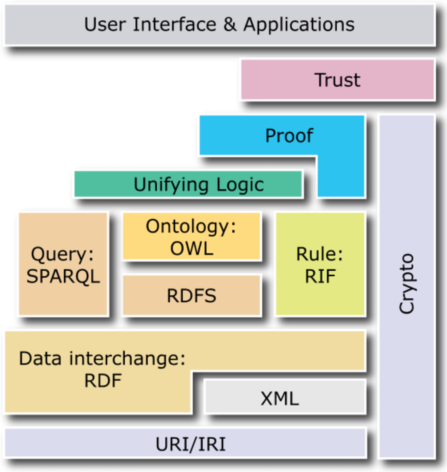


Figure 3.3: The Semantic Web Model (W3C 2013)

3.2.2 Semantic Modelling

MKP Chapter 2:
semantic models
- help people communicate about a fact or situation in the world
- explain and make predictions about the world
- mediate among multiple viewpoints and allow to explore commonalities as well as differences

1. human communication and modelling:
- helps people to coordinate their understanding collaboratively
- knowledge will be gathered, organized, tagged and shared
- when building models in natural human language they are usually open for interpretation of the meaning (e.g. laws)
- interpretation of the text depends on time and context of use -¿ informal model
- the success of informal models can be measured as degree of people supporting the intended purpose
- tagging systems provide an informal organisation to a large body of heterogenous information
- in addition: models can have different layers with an increasing degree of formality (e.g. in the sector of regulations and laws there are regional, national as well as international laws with different degree of formality)

- informal models might be fitting their purpose in the context of their creation, but might need additional layers of models when their usage get beyond that original context to represent the shared meaning

2. explanations and predictions:
- help individuals to draw their own conclusions based on the information received
- especially useful in “interpretive situations” -¿ something is not set in stone
- explanation plays a crucial role in the “understanding” of a situation; if someone can “explain” it, they usually understood it
- in the Semantic Web explanation might help reuse the whole or parts of an existing model
- prediction is closely related to explanation; if a model offer an explanation for a certain situation, it can also be used to make predictions
- that resembles the fundamental of the scientific method (falsification)
- explanation and prediction require a more formal models than used for human communication (see above)
- usually they are build up from objective statements that are used to describe principles and rules (aka formalism)
- these models can also be used to make predictions
- they allow to evaluate the validity of a model and its applicability to a given situation
- in opposite to human communication formalism doesn’t need extra layers of explanations
- in the Semantic Web there are certain standards (a formalism) for modelling explanations
- these techniques can also be used to validate proofs and make predictions (aka inference)

3. Mediating Variability:
- goes hand in hand with AAA principle of the Semantic Web
- usually one decides for a specific viewpoint based on the information from trusted authorities
- informal approach: let every opinion stay side-by-side and let the consumer choose which one to follow
- in this scenario the notion depends on the readers interpretation (as is also common in the Web of information)
- can be modelled in an OOP sense with classes and a hierarchy between them (the higher the more general, the lower the more specific)
- works well for known categories of entities (aka taxonomies)

- any model can also be build up from contributions from multiple sources
- usually seen as layers from different sources
- combination of all layers into a complete model
- a simple merge operation on the layers is easy, but might also introduce inconsistencies of viewpoints into the model
- when two or more viewpoints come together on the Semantic Web there will be an overlap of information
- this will result in disagreements and confusions in the beginning before there will be synergy, cooperation and collaboration
- essence of the Semantic Web: provide an infrastructure that supports AAA and help the community to work through the resulting information chaos to come up with a shared meaning

4. Level of expressivity:
- different people contribute information on different levels of expressivity
- each level might be sufficient to answer specific questions while leaving out unnecessary (sometimes confusing and complex) details
- as of this each level has its purpose!
- also on the Semantic Web there are tools for different levels of expressivity, from the least to the most expressive:
1) RDF: foundation for making statements
2) RDFS: basic notion of classes, hierarchies and relationships
3) RDFS+: subset of OWL, more expressive as RDFS, less complex than OWL, but no standard yet. tries to solve some issues with RDFS for industry use
4) OWL: express logic on the Semantic Web like contraints between classes, entities and relationships

- in the context of the Semantic Web modelling is an ongoing process with some well-structured knowledge and some new, unstructured information coming in at the same point in time

3.2.3 Resource Description Language

MIT Chapter 2:
what is needed to exchange information?
1. syntax: how to serialize the data?

2. data model: how to structure and organize the data?
3. semantics: how to interpret the data?

HTML is made for rendering information on screen and for human consumption
RDF brings a flexible data model to the Web:
- basic building block is a **triple** of *entity* - *attribute* - *value* also known as statement (could also be expressed as *subject* - *predicate* - *object*)
RDFS describes the vocabulary that is available

so:
1. syntax: Turtle, RDFa, RDF-XML or JSON-LD
2. data model: RDF
3. semantics: RDFS

foundational elements are:
- resources (aka just a "thing" of interest identified by an URI or URL depending on its accessibility)
- properties (specify the relations between resources, also identified by URIs)
- statements (assign a value to a 'resource-property' relation, value could be another resource or a literal)
- graphs (RDF is a graph-centered data model, could be distributed, Web of Data / Linked Data approaches)

linked data principles:
- use URIs as name for things
- use HTTP URLs so ppl. can look up those things on the Web
- if they do so, provide useful information (HTML and/or RDF, content and/or meta data)
- include links to other URLs so they can discover more/related things

named graph:
- can be used to point to specific statements or (sub-)graphs
- alternative: reification via an auxiliary object

Turtle: Terse RDF triple language
- <subject incl. URI><predicate incl. URI><object incl. URI>.
- literals will be expressed as "value"ˆˆ<XML schema data type>and supports *string, integer, decimal, dates, . . .*

- URIs can be prefixed: @prefix: <URI>
- repetition: ‘;’ repeats the subject from previous statement, ‘,’ repeats subject and predicate from previous statement
- named graphs in Turtle via Trig extension:
[...] <predicate incl. URI> [...]

`sample.ttl:`

```
@prefix ns1: <URI>
@prefix ns2: <URI>
@prefix ns3: <URI>

ns1:subject ns2:predicate ns3:object .
```

RDF/XML: RDF represented in XML format
- RDF namespace and root node
- subjects in ‘RDF:description’ node containing ‘RDF:about’ attribute with URI
- predicates and objects are child elements of subject node
- use XML namespeaces for URI of nodes

`sample.xml:`

```
<rdf:Description rdf:about="<subject incl. URI>">
  <ns2:predicate rdf:resource="<object incl. URI>" />
</rdf:Description>
```

RDFa: mixin RDF meta-data into HTML
- ‘about’ attribute on <span>or <div>in HTML
- ‘property’ attribute for literal value assignment
- ‘rel’ and ‘resource’ attributes for non-literals
- use XML namespaces for URI of data nodes
- put ‘[]’ around subject and object notations

`sample.html:`

```
<div about="[ns1:subject]">
  <span rel="ns2:relation" resource="[ns3:object]">
</div>
```

MKP Chapter 3:
- usually data is provided in tables from a database
- if we wanna split those over multiple servers, we can:
1) simply split the tables on a row-basis; the table needs to have the same layout on all servers
2) simply split the tables on a column-basis; the rows in each column need an unique identifier to match up the results
3) break down the whole table into cells and distribute them across all servers
->cells with facts need an unique identifier for the row as well as the column

- therefore RDF uses a triple of subject - predicate - object
- subject and predicate are using an unique identifier based on URI
- the triple can be visualized as directed graph

- data from multiple sources can be combined into a graph, if it can be figured out, which nodes exist in both distributed graphs
- therefore nodes are prefixed with an URI
- this URI should be an URL if the information can be dereferenced on the World-Wide Web
- usually they are used in combination with qnames, which define abbrevations for full-qualified URIs
e.g. qname <URI>
qname:subject predicate qname:object .
- use camel case for identifiers, no spaces are allowed
- W3C defines some qnames themselves:
- rdf: contains identifiers used in RDF
- rdfs: contains identifiers used in RDFS
- owl: contains identifiers used in OWL

- in any case: if you use URLs for your entities at least provide a Web page with the explanation of them

- use rdf:type to specify the type of a subject or object (e.g. geo:Berlin rdf:type geo:City .)
- use rdf:Property to specify an identifier to be used as a predicate (e.g. geo:latitude rdf:type rdf:Property .)
- the references objects could also be literal objects like numbers, dates and strings (they borrow the data type specifications from the XML standard)

- statements can also refer to other statements; this kind of metadata about statements can include:
1) provenance (who has made the statement)
2) likelyhood (what is the probability of this statement)
3) context (the setting in which the statement is valid)
4) timeframe (the time constraints for this statement)

- explicit reification with the predicates rdf:subject, rdf:predicate, rdf:object; e.g.:
q:n1 rdf:subject geo:Berlin
rdf:predicate geo:size
rdf:object geo:MegaCity .

web:Wikipedia m:says q:n1 .
- this sample just qualifies that a source (here: Wikipedia) has made a certain statement (n1); but does say nothing about the statement itself! it is up to the application to decide whether the source (Wikipedia) can be trusted or not!

- RDF triples can be serialized as:
1) N-Triples
2) Turtle
3) RDF/XML
4) RDFa

- blank nodes are commonly used to express unknown or uncertain entities
- they will be described in turtle within []
- an ordered set of items can be represented in turtle as ()

3.2.4 Web Ontologies

Lightweight approach: RDFS
- is about adding semantics to your RDF documents

Start by:
1. specify the **things** to talk about
differentiate between *objects* (real entities) and *classes* (set of entities)

‘rdf:type’ attribute to assign objects to classes (object = instance of this class)
impose restrictions on the kind of properties used on objects:
- restrictions on values are called ‘range’ restrictions (object can take values of . . .)
- restrictions on property-object relations are called ‘domain’ restrictions (this relation applies to objects of . . .)
2. set up relations between classes (inheritance, composition)
3. define properties (registered globally) and the possible hierarchy relationship between them (global properties means you can extend existing RDFS classes with your own properties easily)

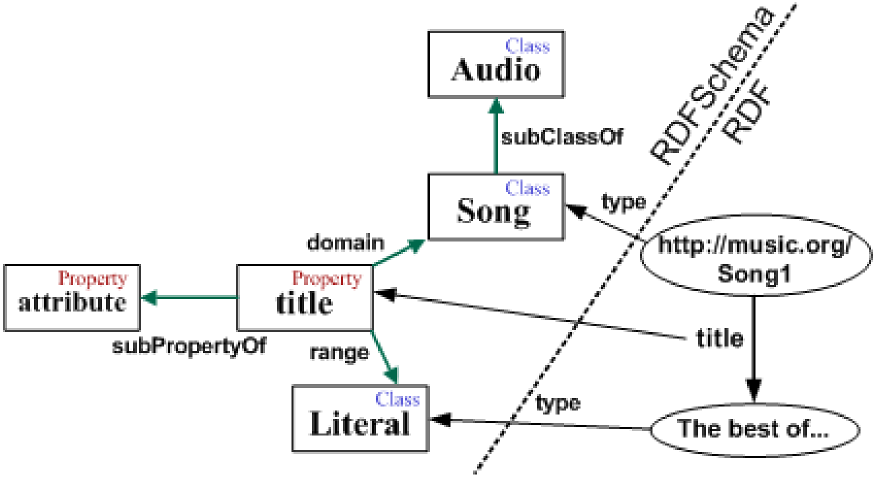


Figure 3.4: RDF Schema sample

RDFS is described in RDF style using:
- core classes like:
- ‘rdfs:Resource’ (all objects/resources)
- ‘rdfs:Class’ (all classes)
- ‘rdfs:Literal’ (all literals)
- ‘rdfs:Property’ (all properties)
- ‘rdfs:Statement’ (all reified statements)
- core properties like:
- ‘rdfs:type’ (specify kind of class)
- ‘rdfs:subClassOf’ (specify inheritance between classes)
- ‘rdfs:subPropertyOf’ (specify inheritance between properties)
- ‘rdfs:domain’ (specify domain restrictions)
- ‘rdfs:range’ (specify range restrictions)

- container classes like:
- ‘rdf:Bag’ (unordered list of entitites)
- ‘rdf:Seq’ (ordered list of entities)
- ‘rdf:Alt’ (list of alternatives/choices)
- ‘rdf:Container’ (superclass for all containers)
- utility classes like:
- ‘rdfs:seeAlso’, ‘rdfs:isDefinedBy’ (links and references to other entities)
- ‘rdfs:Comment’ (comments and notes of entities)
- ‘rdfs:Label’ (human-friendly name of entities)

Missing features in RDFS: . . .

Complex Ontologies in Web Ontology Language (OWL):
. . .

3.2.5 Query Language

SPARQL requires a **triple store** - a database containing RDF documents
is also referred to as a *Graph Store*
data is inserted via Bulk load operation or via SPARQL update statements
SPARQL consist of SPARQL Queries that are send over the SPARQL protocol
Clients sends the queries to an HTTP endpoint
Stores on the public Web incl. dbpedia.org, ckan.org, wikidata.org
SPARQL also works with RDFS
SPARQL has similarities to SQL: - each element in a triple might be replaced with a variable like ‘?varName’ like so:

`sample.sparql`:

```
PREFIX ns1:<URI>
PREFIX ns2:<URI>
PREFIX ns3:<URI>

SELECT ?varName
WHERE {
    ns1:subject ns2:predicate ?varName
}
```

- in the WHERE clause it hosts the graph pattern to match (could be cascaded to go down subgraphs)
- variables can occur at any place in the graph pattern (?subj ?pred ?obj) as select with query everything

LIMIT <n>option at the end for limiting the result set
FILTER (?varName <condition>) in graph pattern can restrict results to match some literal values and supports:
- numbers, dates: <, >, =
- strings: =, regex()

open world assumption: resources on the Web are described in different schematas with various properties using different vocabularies
- UNION option in graph pattern combines different matches
- OPTIONAL option in graph pattern only returns those entities if they are available (otherwise empty)

ASK query checks for the existence of a given graph pattern
CONSTRUCT can be used to retrieve a subgraph from a larger graph, can also be used to translate between different schemas
`sample2.sparql:`

```
PREFIX ns1:<URI>
PREFIX ns2:<URI>
PREFIX ns3:<URI>

CONSTRUCT {
    ?varA ns2:predicate ?varB .
    ?varA ns3:predicate ?literalA .
}
WHERE {
    ?varA ns1:predicate ?varB
}
FILTER ( ?varB > x )
```

- SPARQL can be used to harmonize graphs from different sources
- is also used for basic reasoning ala “if found this, assume that”
- can ease hierarchical queries with * or + on the predicate (SPARQL 1.1)
- can help resolving issues with different entities referring to the same object (MKP

pg. 95)
- Federated Queries can be used to combine information from distinct sources via SPARQL (MKP pg. 110-112)

- inferencing information from existing triples via SPIN (SPARQL Inferencing Notation)
- like in a taxonomy items can be categorized in an hierarchy (MKP pg. 114)
- inference patterns are used in Semantic Web applications (MKP pg. 115)
* subClassOf - type propagation rule
- inferencing could be done at query time or persistently (MKP pg. 120/121)
- inferences can also be helpful when combining information from unknown sources
- inferencing happens on various levels (RDFS, RDFS+, OWL) with an increased set of complex inferencing rules (MKP pg. 122/123)

3.2.6 Agents and Rules

3.3 Peer-to-peer communication

3.3.1 Centralized vs. Decentralized Web Architectures

- in a classical client-server scenario a single server is storing information and distributing it to the clients
- the information is centralized and under control of the provider

- a P2P network considers all nodes equal
- each node can provide information to any other node
- information in a P2P network has to be indexed so that the correct node is queried for it
- the index itself has to be stored somewhere (e.g. on a central server like Napster or in a distributed manner spread over the nodes of the P2P network)

- a P2P system has an high degree of decentralization
- the system is usually self-organizing (adding new or removing members automatically)
- the whole system is usually not controlled by a single organisation and spread over various domains
- it tends to be more resilent to faults and attacks
- can be used for file & data sharing, media streaming, telephony, volunteer computing

and much more

- can be categorized by the degree of centralization into:
1) partly centralized P2P systems (have a dedicated controller node that maintains the set of participating nodes and controls the system)
2) decentralized P2P systems (there are no dedicated nodes that are critical for the system operation)

3.3.2 Initiating a communication session

- depends on the structure of the P2P system - in a partly centralized P2P system new nodes join the network by connecting to the central controller (wellknown IP address)
- in a decentralized P2P system new nodes are expected to obtain via a separate channel the IP address to connect to (usually a bootstrap node that helps to set up the new node)

3.3.3 Finding communication peers

- also known as the overlay network in a P2P system
- can be represented as a directed graph containing the nodes and communication links between them
- can be differentiated between unstructured and structured overlays
- unstructured overlay networks have no constraints for the links between nodes; therefore the network has no particular structure
- structured overlay networks assign an unique identifier from a numeric keyspace to each node; these keys are used to assign certain responsibilities to nodes on the network; as of this routing can be handled more efficiently
- in partly centralized P2P systems the controller is responsible for the overlay formation

- in partly centralized P2P system an object is typically stored at the node that inserted the object
- the central controller holds the information about which objects exist and which nodes hold them

- in unstructured systems the information is typically stored on the nodes that introduces them
- to locate an object a query request is typically broadcasted through the overlay network
- often the scope of the request (e.g. the maximum number of hops from the querying node forward) is limited to reduce the overhead on the system

- in structured systems a distributed index is maintained in the form of a distributed hash table
- this DHT holds the hash value of the (index) key and the address of the node that stores the value

3.3.4 Transmitting Data

3.3.5 Available Protocols

4 Related Works

Related works, that are referenced for this thesis, are:

1. Scenario:
- “Fraud in Non-Cash Transactions: Methods, Tendencies and Threats.” (Sobko 2014)
- “Overview of E-Commerce” (Ankhule & Joshy 2015)
- “A Survey on Fraud Detection Techniques in Ecommerce” (Rana & Baria 2015)
- “A Study on E-Commerce Security Issues and Solutions” (Sen et al. 2015)

2. CSCW:
- “Effects of Sensemaking Translucence on Distributed Collaborative Analysis” (Goyal & Fussell)
- “CSCW and enterprise 2.0 - towards an integrated perspective” (Koch 2008)
- “A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network” (Yang & Chen 2008)
- “Paradox of richness: A cognitive model of media choice” (Robert & Dennis 2005)
- “From The Matrix to a Model of Coordinated Action (MoCA): A Conceptual Framework of and for CSCW” (Lee & Paine 2015)

3. P2P:
- “SWAP: Ontology-based Knowledge Management with Peer-to-Peer Technology.” (Ehrig et al. 2003)
- “RDFPeers: a scalable distributed RDF repository based on a structured peer-to-peer network” (Cai & Frank 2004)
- “P2P networking: An information-sharing alternative” (Parameswaran et al. 2001)
- “Introduction to XMPP protocol and developing online collaboration applications using open source software and libraries” (Ozturk 2010)
- “Peer-to-peer systems” (Rodrigues & Druschel 2010)

4. Semantic Web:
- “Semantic web technologies for the financial domain” (Lara et al. 2007)
- “Gephi: an open source software for exploring and manipulating networks.” (Bastian

et al. 2009)
- “Security ontology: Simulating threats to corporate assets” (Ekelhart et al. 2006)
- “Marvin: Distributed reasoning over large-scale Semantic Web data” (Oren et al. 2009)
- “Applying Semantic Technologies to Fight Online Banking Fraud” (Carvalho et al.)
- “The Semantic Web-Based Collaborative Knowledge Management” (Chao et al. 2012)
- “Open eBusiness Ontology Usage: Investigating Community Implementation of GoodRelations.” (Ashraf et al. 2011)
- “Hexastore: sextuple indexing for semantic web data management” (Weiss et al. 2008)
- “Rule responder: RuleML-based agents for distributed collaboration on the pragmatic web” (Paschke et al. 2007)
- “Rule interchange on the web” (Boley et al. 2007)
- “An owl-based security incident ontology” (Martimiano & Moreira 2005)
- “Data linking for the semantic web” (Scharffe et al. 2011)
- “On enhancing scalability for distributed RDF/S stores” (Tsatsanifos et al. 2011)
- “Integrating agents, ontologies, and semantic web services for collaboration on the semantic web” (Stollberg & Strang 2005)
- “GoodRelations Tools and Applications” (Hepp et al. 2009)
- “Drawing Conclusions from Linked Data on the Web: The EYE Reasoner” (Verborgh & De Roo 2015)
- “Toward a security ontology” (Donner 2003)
- “Schema.org: Evolution of structured data on the web” (Guha et al. 2016)
- “Goodrelations: An ontology for describing products and services offers on the web” (Hepp 2008)
- “The stac (security toolbox: attacks & countermeasures) ontology” (Gyrard et al. 2013)
- “Formalizing information security knowledge” (Fenz & Ekelhart 2009)
- “A functional semantic web architecture” (Gerber et al. 2008)
- “Developing an Ontology of the Cyber Security Domain.” (Obrst et al. 2012)
- “Towards a financial fraud ontology: A legal modelling approach” (Kingston et al. 2004)
- “Complete query answering over horn ontologies using a triple store” (Zhou et al. 2013)

5 Concept and Design of the System

ca. 30 pages

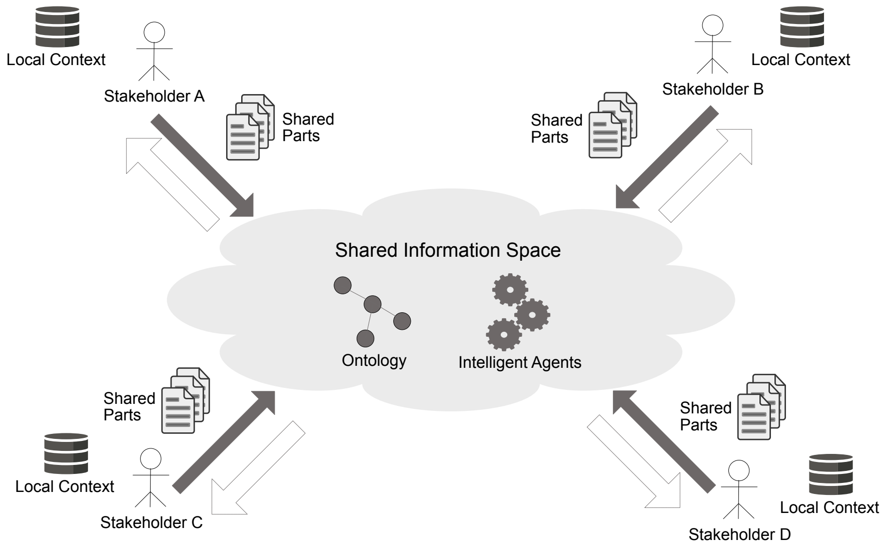


Figure 5.1: System Overview

Based on chapter 3.1 we can conclude:
1. Face-to-Face Meetings: out-of-scope of this thesis
2. Distance Meetings: lack of collaboration support
3. Continuous Tasks: collaboration in teams, but only works when everyone is online
4. Communicate & Collaborate: allows to work on it in a disconnected mode, but increases communication and coordination efforts as well as might lead to synchronisation issues over time

This either leaves us with two options:
1. build a distributed, synchronous collaboration system, in that ppl. can share and work on content at the same time
2. build a distributed, asynchronous collaboration and communication system, in that ppl. can work on things for themselves and get connected together at a certain point in time for synchronising their findings and develop new insights

In the first variant it can be assumed that:
- stakeholders will initiate a collaborative session for a certain case, the collaboration and information sharing efforts end with finishing the case.
- each stakeholder might just work on his part of expertise in the whole knowledge graph (e.g. named subgraphs per stakeholder). these parts could be easily mirrored on the stakeholders environment (no discrepancies with informations from others)
- the whole knowledge graph is only available during the p2p collaboration session, nevertheless results and findings (per stakeholder?) can be synchronized into the named graph of the stakeholder and be analysed offline
- . . .

In the second variant it can be assumed that:
- every stakeholder holds different parts of the whole knowledge graph, even might hold the whole graph on his machine.
- stakeholders can fill out the information offline, they might get together at irregular intervals to synchronise their efforts and come up with new knowledge graph entries based on the work of the others
- during the synchronisation process there might come up discrepancies due to the different understandings of the stakeholders for a certain aspect of the knowledge graph
- there might also be different findings or result, even contradictory statements, based on the different progress of each stakeholder on the knowledge graph
- . . .

Based on chapter 3.2.2 we might come up with a design of the shared information space that looks like this:

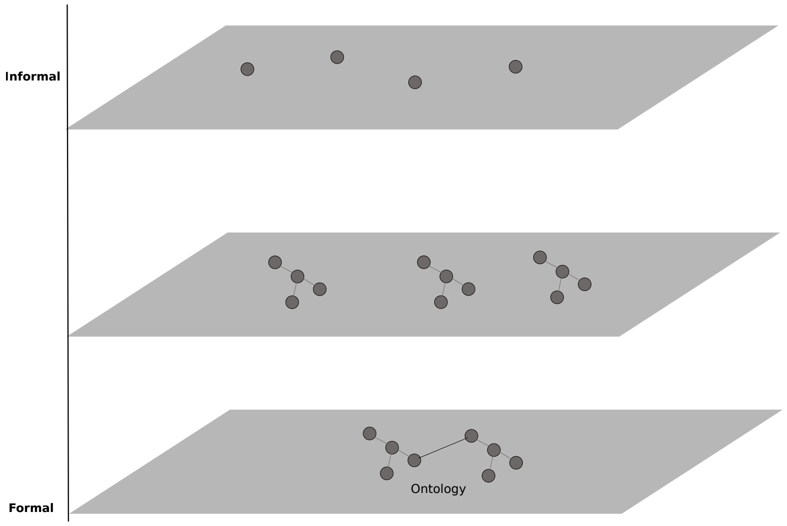


Figure 5.2: Shared Information Space

- it uses different layers to connect domain specific knowledge
- these layers are ordered based on formalism and models available
- the more informal a layer is the more collaboration is required to come up with a shared understanding / the common sense and connect it with the layer below
- layers can be easily turned on / off in the application to focus on a specific aspect of the investigation

Based on chapter 3.3 we can either do:
- a partially centralized P2P system: having the supernodes that hold the information about the participating parties as well as the whole graph for analyses at trustworthy parties like the BKA

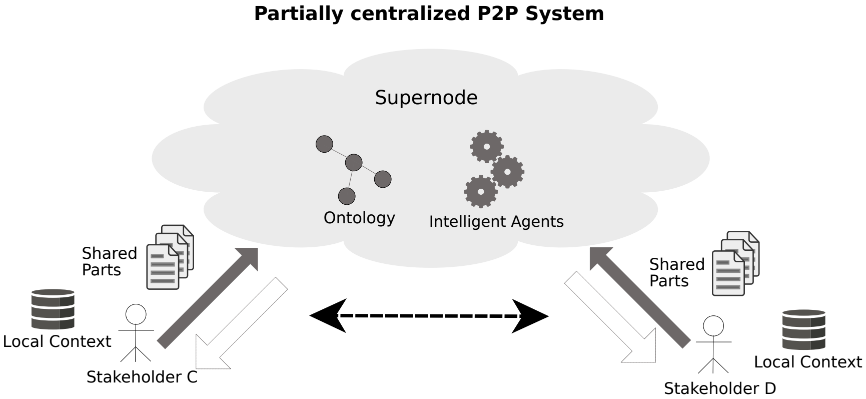


Figure 5.3: A centralized P2P system design

- a decentralized P2P system in which every node is equal: information is spread over the various nodes; each party holds the information from the local context (domain specific) and provide an access point incl. API (e.g. SPARQL) for publicly available (shared) content
- in this case the intelligent agents can use a mechanism like MapReduce to send the analytics algorithm to the endpoints of the participating parties and asking for information needed to draw conclusions locally

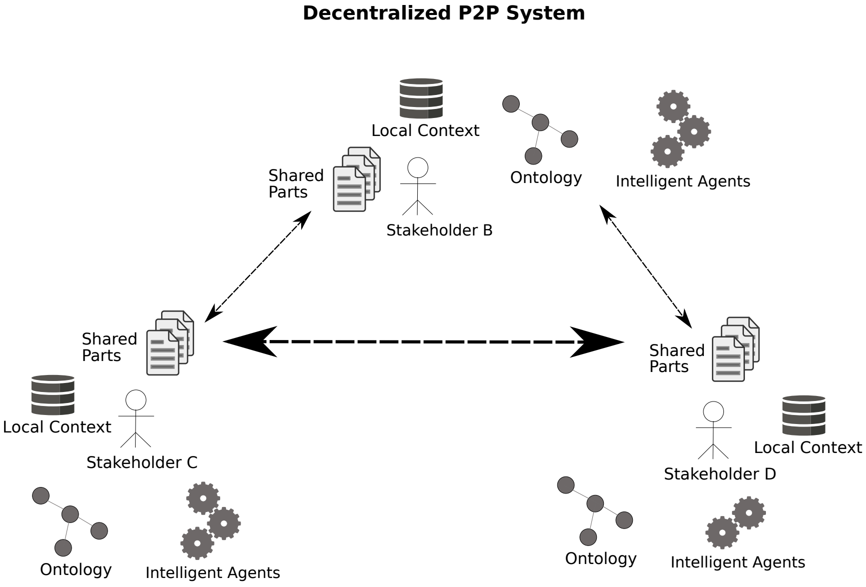


Figure 5.4: A decentralized P2P system design

Reusable Semantic Web resources on the Web:
- Ontologies:

1) GoodRelations: describing E-commerce shop transactions

2) Friend-of-a-Friend: describing social relationships between people

3) DublinCore: describing artefacts found on the Web 2.0 (author, date, license, ...)

4) Web Of Trust: describing trust related entities for Web documents (PGP infrastructure)

- Sites:

1) DAML: finding ontologies or taxonomies for various domains

2) UMBEL: integrating and mapping technologies and vocabularies for ontologies

3) Wikidata: a free knowledge base representing information from Wikipedia as RDF

6 Conclusion and Future Work

ca. 10 pages

List of Figures

List of Tables

Glossary

B2B	Business-To-Business.
B2C	Business-To-Consumer.
C2B	Consumer-To-Business.
C2C	Consumer-To-Consumer.
CSCW	computer-supported cooperative work.
CSP	Cloud Service Provider / Hosting Service.
IP	Internet Protocol.
ISP	Internet Service Provider.
ISV	Independent Software Vendor.
LSP	Logistic Service Provider.
OWL	Web Ontology Language.
P2P	Peer-To-Peer.
PSP	Payment Service Provider.
RDF	Resource Description Framework.
SPARQL	SPARQL Protocol and RDF Query Language.
URL	Uniform Resource Locator.
WebRTC	Web Real-Time Communication.
XMPP	Extensible Messaging and Presence Protocol.

Bibliography

Ankhule & Joshy 2015
Ankhule, Gayatri R.; Joshy, MR: Overview of E-Commerce. In: *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)* (2015), pages 196

Ashraf et al. 2011
Ashraf, Jamshaid; Cyganiak, Richard; O'Riain, Seán; Hadzic, Maja: *Open eBusiness Ontology Usage: Investigating Community Implementation of GoodRelations.* In: *LDOW*, 2011

Bastian et al. 2009
Bastian, Mathieu; Heymann, Sebastien; Jacomy, Mathieu et al.: Gephi: an open source software for exploring and manipulating networks. In: *ICWSM* 8 (2009), pages 361–362

Boley et al. 2007
Boley, Harold; Kifer, Michael; Pătrânjan, Paula-Lavinia; Polleres, Axel: Rule interchange on the web. In: *Reasoning Web.* Springer, 2007, pages 269–309

Brachmann 2015
Brachmann, Steve: *In the face of growing e-commerce fraud, many merchants not prepared for holidays - IPWatchdog.com |patents & patent law.* http://www.ipwatchdog.com/2015/11/22/growing-e-commerce-fraud-merchants-not-prepared-for-holidays/id=63271/. Version: 11 2015

Business Wire 2015
Business Wire: Global card fraud losses reach $16.31 Billion — will exceed $35 Billion in 2020 according to the Nilson report. In: *Business Wire* (2015), 08. http://www.marketwatch.com/story/global-card-fraud-losses-reach-1631-billion-will-exceed-35-billion-in-2020-accor

Cai & Frank 2004
Cai, Min; Frank, Martin: *RDFPeers: a scalable distributed RDF repository based on a structured peer-to-peer network.* In: *Proceedings of the 13th international conference on World Wide Web* ACM, 2004, pages 650–657

Captain 2015
Captain, Sean: These are the mobile sites leaking credit card data for up to 500, 000 people A day. In: *Fast Company* (2015), 12. http://www.fastcompany.com/3054411/these-are-the-faulty-apps-leaking-credit-card-data-for-up-to-500000-people-a-day

Carvalho et al.
Carvalho, Rodrigo; Goldsmith, Michael; Creese, Sadie; Police, Brazilian F.: Applying Semantic Technologies to Fight Online Banking Fraud.

Chao et al. 2012
Chao, Lemen; Xing, Chunxiao; Zhang, Yong: *The Semantic Web-Based Collaborative Knowledge Management.* INTECH Open Access Publisher, 2012

Consumer Action 2009
Consumer Action: Questions and answers about credit card fraud A Q & consumer aCtion A consumer action publication. Version: 2009. `http://www.consumer-action.org/downloads/english/Chase_CC_Fraud_Leaders.pdf`. http://www.consumer-action.org/downloads/english/Chase_CC_Fraud_Leaders.pdf, 2009. – Forschungsbericht

Donner 2003
Donner, Marc: Toward a security ontology. In: *IEEE Security & Privacy* (2003), Nr. 3, pages 6–7

DSS 2014
DSS, PCI: *Payment Card Industry Data Security Standards.* 2014

Ehrig et al. 2003
Ehrig, Marc; Tempich, Christoph; Broekstra, Jeen; Van Harmelen, Frank; Sabou, Marta; Siebes, Ronny; Staab, Steffen; Stuckenschmidt, Heiner: *SWAP: Ontology-based Knowledge Management with Peer-to-Peer Technology.* In: *Wissensmanagement*, 2003, pages 17–20

Ekelhart et al. 2006
Ekelhart, Andreas; Fenz, Stefan; Klemen, Markus D.; Weippl, Edgar R.: *Security ontology: Simulating threats to corporate assets.* Springer, 2006

Fenz & Ekelhart 2009
Fenz, Stefan; Ekelhart, Andreas: *Formalizing information security knowledge.* In: *Proceedings of the 4th international Symposium on information, Computer, and Communications Security* ACM, 2009, pages 183–194

Gerber et al. 2008
Gerber, Aurona; Merwe, Alta Van d.; Barnard, Andries: *A functional semantic web architecture.* Springer, 2008

Google Patents
`https://patents.google.com/?q=credit+card+fraud+prevention&after=20150101`

Goyal & Fussell
Goyal, Nitesh; Fussell, Susan R.: Effects of Sensemaking Translucence on Distributed Collaborative Analysis.

Guha et al. 2016

Guha, RV; Brickley, Dan; Macbeth, Steve: Schema. org: Evolution of structured data on the web. In: *Communications of the ACM* 59 (2016), Nr. 2, pages 44–51

Gyrard et al. 2013

Gyrard, Amelie; Bonnet, Christian; Boudaoud, Karima: *The stac (security toolbox: attacks & countermeasures) ontology.* In: *Proceedings of the 22nd international conference on World Wide Web companion* International World Wide Web Conferences Steering Committee, 2013, pages 165–166

Hepp 2008

Hepp, Martin: Goodrelations: An ontology for describing products and services offers on the web. In: *Knowledge Engineering: Practice and Patterns.* Springer, 2008, pages 329–346

Hepp et al. 2009

Hepp, Martin; Radinger, Andreas; Wechselberger, Andreas; Stolz, Alex; Bingel, Daniel; Irmscher, Thomas; Mattern, Mark; Ostheim, Tobias: *GoodRelations Tools and Applications.* In: *Poster and Demo Proceedings of the 8th International Semantic Web Conference (ISWC 2009), Washington, DC, USA*, 2009

Holmes 2015

Holmes, Tamara E.: *Credit card fraud and ID theft statistics.* `http://www.creditcards.com/credit-card-news/credit-card-security-id-theft-fraud-statistics-1276.php`. Version: 09 2015

Kingston et al. 2004

Kingston, John; Schafer, Burkhard; Vandenberghe, Wim: Towards a financial fraud ontology: A legal modelling approach. In: *Artificial Intelligence and Law* 12 (2004), Nr. 4, pages 419–446

Koch 2008

Koch, Michael: *CSCW and enterprise 2.0 - towards an integrated perspective.* In: *BLED 2008 Proceedings*, 2008

Lara et al. 2007

Lara, Rubén; Cantador, Iván; Castells, Pablo: Semantic web technologies for the financial domain. In: *The Semantic Web.* Springer, 2007, pages 41–74

Lee & Paine 2015

Lee, Charlotte P.; Paine, Drew: *From The Matrix to a Model of Coordinated Action (MoCA): A Conceptual Framework of and for CSCW.* In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* ACM, 2015, pages 179–194

Lewis 2015
Lewis, Len: *More vulnerable than ever?* https://nrf.com/news/more-vulnerable-ever. Version: 12 2015

Martimiano & Moreira 2005
Martimiano, AFM; Moreira, ES: *An owl-based security incident ontology.* In: *Proceedings of the Eighth International Protege Conference*, 2005, pages 43–44

Obrst et al. 2012
Obrst, Leo; Chase, Penny; Markeloff, Richard: *Developing an Ontology of the Cyber Security Domain.* In: *STIDS*, 2012, pages 49–56

Oren et al. 2009
Oren, Eyal; Kotoulas, Spyros; Anadiotis, George; Siebes, Ronny; Teije, Annette ten; Harmelen, Frank van: Marvin: Distributed reasoning over large-scale Semantic Web data. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 7 (2009), Nr. 4, pages 305–316

Ozturk 2010
Ozturk, Ozgur: *Introduction to XMPP protocol and developing online collaboration applications using open source software and libraries.* In: *Collaborative Technologies and Systems (CTS), 2010 International Symposium on* IEEE, 2010, pages 21–25

Parameswaran et al. 2001
Parameswaran, Manoj; Susarla, Anjana; Whinston, Andrew B.: P2P networking: An information-sharing alternative. In: *Computer* (2001), Nr. 7, pages 31–38

Paschke et al. 2007
Paschke, Adrian; Boley, Harold; Kozlenkov, Alexander; Craig, Benjamin: Rule responder: RuleML-based agents for distributed collaboration on the pragmatic web. In: *(RuleML-2008).* (2007)

PYMNTS 2016
PYMNTS: *Hackers and their fraud attack methods.* http://www.pymnts.com/fraud-prevention/2016/benchmarking-hackers-and-their-attack-methods. Version: 02 2016

Rampton 2015
Rampton, John: How online fraud is a growing trend. In: *Forbes* (2015), 04. http://www.forbes.com/sites/johnrampton/2015/04/14/how-online-fraud-is-a-growing-trend/#16ffc0ec349f

Rana & Baria 2015
Rana, Priya J.; Baria, Jwalant: A Survey on Fraud Detection Techniques in Ecommerce. In: *International Journal of Computer Applications* 113 (2015), Nr. 14

Reuters 2015

Reuters: *Fraud rates on online transactions seen up during holidays: Study.* http://www.reuters.com/article/us-retail-fraud-idUSKCN0T611T20151117?feedType=RSS&feedName=technologyNews. Version: 11 2015

Rice 1992

Rice, Ronald E.: Task Analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. In: *Organization Science* 3 (1992), 11, Nr. 4, pages 475–500. http://dx.doi.org/10.1287/orsc.3.4.475. – DOI 10.1287/orsc.3.4.475. – ISSN 1047–7039

Robert & Dennis 2005

Robert, Lionel P.; Dennis, Alan R.: Paradox of richness: A cognitive model of media choice. In: *Professional Communication, IEEE Transactions on* 48 (2005), Nr. 1, pages 10–21

Rodrigues & Druschel 2010

Rodrigues, Rodrigo; Druschel, Peter: Peer-to-peer systems. In: *Communications of the ACM* 53 (2010), Nr. 10, pages 72–82

Scharffe et al. 2011

Scharffe, François; Ferrara, Alfio; Nikolov, Andriy: Data linking for the semantic web. In: *International Journal on Semantic Web and Information Systems* 7 (2011), Nr. 3, pages 46–76

Sen et al. 2015

Sen, Pritikana; Ahmed, Rustam A.; Islam, Md R.: A Study on E-Commerce Security Issues and Solutions. (2015)

Sobko 2014

Sobko, Oleg V.: Fraud in Non-Cash Transactions: Methods, Tendencies and Threats. In: *World Applied Sciences Journal* 29 (2014), Nr. 6, pages 774–778

Staab & Stuckenschmidt 2006

Staab, Steffen (Hrsg.); Stuckenschmidt, Heiner (Hrsg.): *Semantic web and peer-to-peer.* Springer Science + Business Media, 2006. http://dx.doi.org/10.1007/3-540-28347-1. http://dx.doi.org/10.1007/3-540-28347-1. – ISBN 9783540283461

Stollberg & Strang 2005

Stollberg, Michael; Strang, Thomas: *Integrating agents, ontologies, and semantic web services for collaboration on the semantic web.* In: *Proc. of the First International Symposium on Agents and the Semantic Web, AAAI Fall Symposium Series Arlington, Virginia*, 2005

Tsatsanifos et al. 2011

Tsatsanifos, George; Sacharidis, Dimitris; Sellis, Timos: *On enhancing scalability for distributed RDF/S stores.* In: *Proceedings of the 14th International Conference on Extending Database Technology* ACM, 2011, pages 141–152

Verborgh & De Roo 2015

Verborgh, Ruben; De Roo, Jos: Drawing Conclusions from Linked Data on the Web: The EYE Reasoner. In: *IEEE Software* (2015), Nr. 3, pages 23–27

Visa Europe 2014

Visa Europe: *Processing e-commerce payments.* `https://www.visaeurope.com/media/images/processing%20e-commerce%20payments%20guide-73-17337.pdf`. Version: 08 2014

W3C 2013

W3C: *W3C semantic web activity.* `https://www.w3.org/2001/sw/`. Version: 06 2013

Weiss et al. 2008

Weiss, Cathrin; Karras, Panagiotis; Bernstein, Abraham: Hexastore: sextuple indexing for semantic web data management. In: *Proceedings of the VLDB Endowment* 1 (2008), Nr. 1, pages 1008–1019

Yang & Chen 2008

Yang, Stephen J.; Chen, Irene Y.: A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network. In: *International Journal of Human-Computer Studies* 66 (2008), Nr. 1, pages 36–50

Zhou et al. 2013

Zhou, Yujiao; Nenov, Yavor; Grau, Bernardo C.; Horrocks, Ian: Complete query answering over horn ontologies using a triple store. In: *The Semantic Web–ISWC 2013.* Springer, 2013, pages 720–736

Declaration in lieu of oath

I hereby declare that this master thesis was independently composed and authored by myself.

All content and ideas drawn directly or indirectly from external sources are indicated as such. All sources and materials that have been used are referred to in this thesis.

The thesis has not been submitted to any other examining body and has not been published.

Place, date and signature of student
Andreas Gerlach

Appendix