Mathematical tools for Digital Forensics

Jesús Medina

Action Chair DigForASP – CA17124
Department of Mathematics. University of Cádiz, Spain
Email: jesus.medina@uca.es

UCA Universidad de Cádiz
Departamento de Matemáticas

Krakow, Poland
March 29th, 2019

DIGforASP

Outline

Galois connections

- They are inherent in human thinking wherever logical reasoning about certain hierarchical structures is involved.
- Provide the order-preserving passage between two worlds of our imagination.
- The "hierarchies" in the two opposite worlds are reversed or transported when passing to the other world, and going forth and back becomes a stationary process when iterated.

Example

In classical Galois theory, properties of permutation groups are used to study field extensions.

These operators were used before É. Galois

Constructions and computations of different mathematicians in Centuries XVIII and XIX

- Joseph-Louis Lagrange (1736-1813)
- Niels Henrik Abel (1802-1829)
- Évariste Galois (1811-1832)
- Richard Dedekind (1831-1916)

The current name and the formal definition was introduced in 20th century, which was given due to these operators are a generalization of the ones used in Galois theory.

Formal definition

Formal Galois connection definitions

- Garrett Birkhoff (1911-1996). Related to power sets – 1940.
- Øystein Ore (1899-1968) and Jürgen Schmidt. Standard definition – 1944-1953.
- H. Herrlich and M. Hušek. Categorical notions – 1990.

A toy example

Under what conditions is it possible to solve $-x^2 + 2bx + c = \square$*?*

The mathematician Egipt Abū Kāmil study this equation around 880 b. C.

$$
\begin{array}{lll}
-x^2 + 2bx + c > 0 & \text{iff} & c > x^2 - 2bx \\
& \text{iff} & c + b^2 > x^2 - 2bx + b^2 \\
& \text{iff} & c + b^2 > (x - b)^2 \\
& \text{then} & c + b^2 > 0 \\
& \text{iff} & -c < b^2
\end{array}
$$

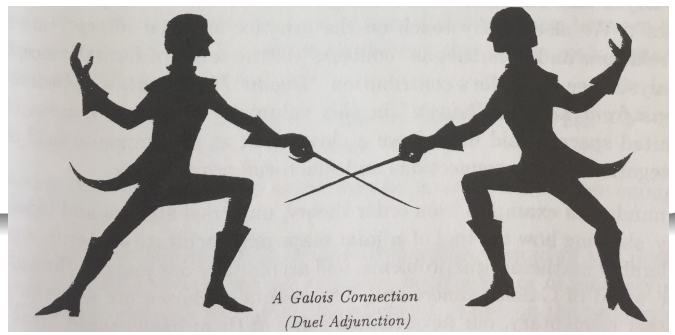
"then" can be an equivalence, when x is almost b.

Galois connection

Galois connection

Let $(P_1, \leq_1)$ and $(P_2, \leq_2)$ be posets, and $^{\downarrow}\colon P_1 \to P_2$, $^{\uparrow}\colon P_2 \to P_1$ mappings, the pair $(^{\uparrow}, ^{\downarrow})$ forms a *Galois connection* between P_1 and P_2 if and only if:

1. $^{\uparrow}$ and $^{\downarrow}$ are order-reversing.
2. $x \leq_1 x^{\downarrow\uparrow}$ for all $x \in P_1$.
3. $y \leq_2 y^{\uparrow\downarrow}$ for all $y \in P_2$.

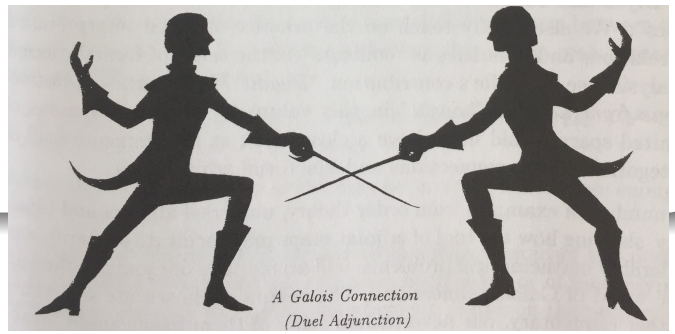


Galois connection

Galois connection

Let $(P_1, \leq_1)$ and $(P_2, \leq_2)$ be posets, and $^{\downarrow}\colon P_1 \to P_2$, $^{\uparrow}\colon P_2 \to P_1$ mappings, the pair $(^{\uparrow}, ^{\downarrow})$ forms a *Galois connection* between P_1 and P_2 if and only if:

1. $^{\uparrow}$ and $^{\downarrow}$ are order-reversing.
2. $x \leq_1 x^{\downarrow\uparrow}$ for all $x \in P_1$.
3. $y \leq_2 y^{\uparrow\downarrow}$ for all $y \in P_2$.



Equivalently

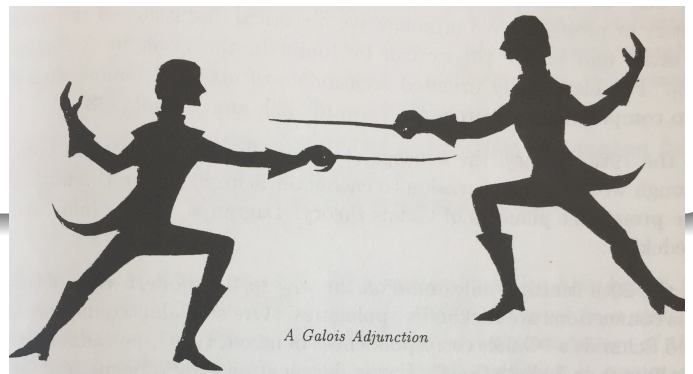
$$x \leq_1 y^{\uparrow} \quad \text{iff} \quad y \leq_2 x^{\downarrow}$$

Adjunction

Adjunction or isotone Galois connection

Let $(P_1, \leq_1)$ and $(P_2, \leq_2)$ be posets, and $^{\downarrow}\colon P_1 \to P_2$, $^{\uparrow}\colon P_2 \to P_1$ mappings, the pair $(^{\uparrow}, ^{\downarrow})$ forms an *isotone Galois connection* between P_1 and P_2 if and only if:

1. $^{\uparrow}$ and $^{\downarrow}$ are order-presersing.
2. $x \leq_1 x^{\downarrow\uparrow}$ for all $x \in P_1$.
3. $y^{\uparrow\downarrow} \leq_2 y$ for all $y \in P_2$.

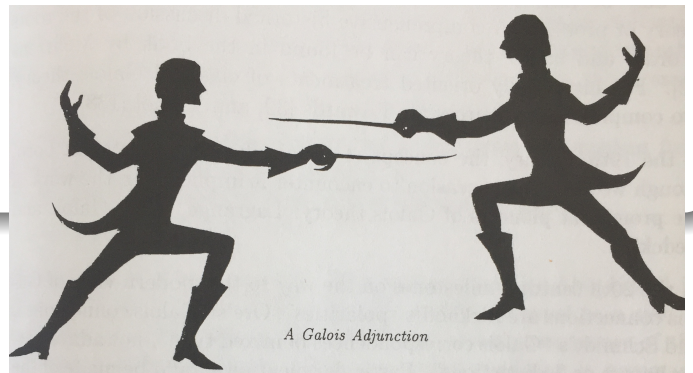


Adjunction

Adjunction or isotone Galois connection

Let $(P_1, \leq_1)$ and $(P_2, \leq_2)$ be posets, and $^{\downarrow}\colon P_1 \to P_2$, $^{\uparrow}\colon P_2 \to P_1$ mappings, the pair $(^{\uparrow}, ^{\downarrow})$ forms an *isotone Galois connection* between P_1 and P_2 if and only if:

1. $^{\uparrow}$ and $^{\downarrow}$ are order-presersing.
2. $x \leq_1 x^{\downarrow\uparrow}$ for all $x \in P_1$.
3. $y^{\uparrow\downarrow} \leq_2 y$ for all $y \in P_2$.



Equivalently

$$y^{\uparrow} \leq_1 x \quad \text{iff} \quad y \leq_2 x^{\downarrow}$$

Useful operators in different theories in the last three centuries

In universal algebra, geometry, topology, logic, category theory, etc.

- Theory of polinomial equations (Lagrange, Galois)
- Modern Galois theory (Dedekind, Artin)
- Origins of lattice theory (Dedekind, Schröder)
- Polarities and lattice-theoretical aspects (Birkhoff)
- Order-theorical Galois connections (Ore)
- Logical calculus (Boole, Pierce, Schröder)
- Residuation theory (Kroll, Ward, Dilworth)

Mathematical tools studied by our research group M·CIS

M·CIS mathematical tools

- Fuzzy sets.
- Fuzzy logic. Fuzzy Logic Programming.
- Fuzzy Formal Concept Analysis.
- Fuzzy Rough Sets.
- Fuzzy Relations Equations.
- Tools for the extraction, manipulation and prediction of information in databases.
- Linguistic description of data and automatic generation of natural language.

Mathematics for Computational Intelligence Systems (M·CIS)

Lab Head

- Jesús Medina. jesus.medina@uca.es

Lab Members

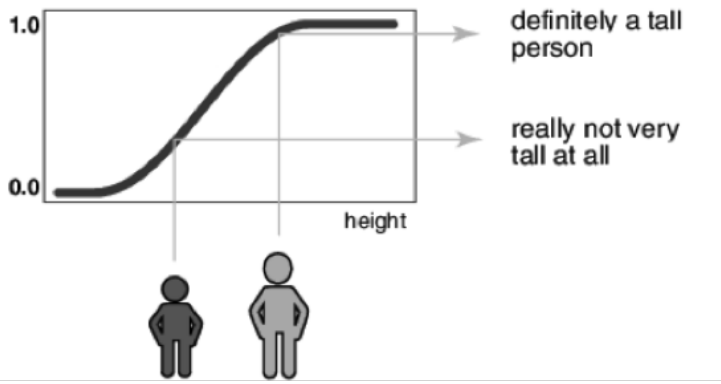
- M. Eugenia Cornejo.
- J. Carlos Díaz-Moreno.
- Eloísa Ramírez-Poussa.
- J. Rafael Rodríguez-Galván.
- Clemente Rubio-Manzano.

In training

- M. José Benítez-Caballero
- David Lobo
- Roberto G. Aragón

Fuzzy Sets. Fuzzy Logic

- An inherent condition to real problems is containing innacurate or incomplete information.
- Fuzzy sets and fuzzy logic allow modeling and solving real problems thanks to their capability to represent inaccurate or deficient information, in a quantitative or qualitative way.
- Fuzzy logic is a technique increasingly recognized due to its successful applications and implementations.



Influence of Fuzzy Logic on Society

Fuzzy logic was introduced by Lotfi A. Zadeh in 1965

70s: Mamdani designed a fuzzy control system to regulate of a steam engine.

80s: Control of a water treatment plant and braking control of the subway (Fuji Electric & Hitachi).

90s: Technology based on fuzzy logic appeared in a large number of home appliances

Influence of Fuzzy Logic on Society

Fuzzy logic has been applied to fields as diverse as:

- Energy efficiency
- Renewable energy
- Security
- Classification and recognition of manuscripts
- Simulation human behavior
- Medical diagnostic
- Prediction of diseases
- Space Research
- Etc.

Advantages of these mathematical tools

M·CIS mathematical tools features

- Reliability
- Robustness in the method
- Flexibility
- White box. NO black box
- You can perform a traceability
- These can be complemented by other interesting machine learning tools.

Outstanding contracts with companies

- Optimization in collection and treatment of agricultural data using fuzzy techniques.
- Application of fuzzy techniques to optimize crop production.
- BIG DATA project in the CBC. Predictive analysis of production processes. Aeronautical branch.
- Development of advanced manufacturing technologies in the aeronautical industry.
- Efficient energy. Photovoltaic facilities

BIntelMAS

BIntelMAS is a system we are developing and it implements some of the techniques we work, obtaining for example

- Minimal units of information extracted from the knowledge system and its hierarchical organization
- Select the most representative variables, removing unnecessary variables
- Rules that model the knowledge system
- Diagnosis and prediction of events
- Learning of the system after the knowledge of the expert

Application of fuzzy techniques to optimize crop production

The following table shows the amounts of fertilizer used in a crop production and the benefit obtained, per year

	N	P_2O_5	K_2O	Eur
06-07	0.78	0.36	0.71	0.64
07-08	1.00	1.00	1.00	1.00
08-09	0.52	0.68	0.65	0.00
09-10	0.55	0.8	0.72	0.52
10-11	0.39	0.66	0.48	0.34
12-13	0.00	0.00	0.00	0.01

```
K20/0.50 N/0.75 P205/0.25  =>  eur/0.50  (0.33, 1.00)
```

Application of fuzzy techniques to optimize crop production

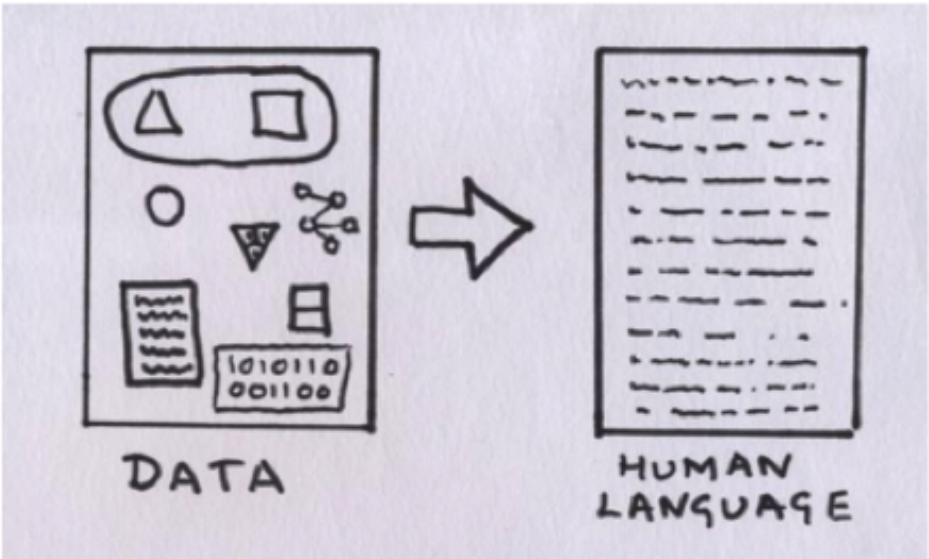
```
K20/0.50 N/0.75 P205/0.25  =>  eur/0.50  (0.33, 1.00)
```

"Always that we throw in the crop at least half (0.50) of the average consumption of K2O, three quarters (0.75) of N and a quarter (0.25) of P2O5, we will obtain at least half (0.5) of the benefit."

Decrease the costs and increase the benefits

Natural language generation

- The generated rules can be difficult to interpret.
- We make the information more accountable to any user, by using natural language generation.



Some advantages of the linguistic description and the generation of natural language: Utility, Time saving, Clarity, Simplicity.

Decision rules in Digital Forensics

Decision rules can be useful to extract knowledge with the goal of modelling behavior patterns in unlawful transactions. For example, in order to detect and prevent fraud related to transactional credit card databases.

Male/1.0 /young/0.8 recent card/0.7 credit card/1.0 big city/0.9 => fraud/0.7

Weighted decision rules in Digital Forensics

We can also define over the rules different measures

- The degree of relevance of each rule: support.
- The the degree of confidence of each one: confidence, certainty factor, ...

From the analysis of these rules can be discovered information related to

- The type of population is most likely to be affected by this credit card fraud.
- Kind of businesses more prone to fraud.
- Areas more prone to fraud.

Example of Fuzzy Logic in Digital Forensics

A detective is investigating a crime. He has a list of suspects stored in a database with the following variables: identification, sex, age, eye_color, hair_colour, height. The witness describes to the suspect as: "a tall person, young with light eyes"

```
suspect(id1,male,age#58,black,dark,height#180).
suspect(id2,female,age#34,green,dark,height#170).
suspect(id3,male,age#20,blue,brown,height#175).
```

If the witness describes

```
?.-(X,_,young,light,_,tall).

Nosuspectsarefound
```

Fuzzy logic is useful here to establish relationships between

- Similar concepts: green and blue are similar to light.
- Scalar magnitude and concepts, the well-known linguistic variables.

Example of Fuzzy Logic in Digital Forensics

By using fuzzy logic programming and fuzzy unification techniques

```
//e.g., obtained from a thesaurus
light~green = 1.0
light~blue = 1.0
light~black = 0.1

//e.g., obtained from a imaginary linguistic variables
tall~height#180 = 0.9
tall~height#175 = 0.6
tall~height#170 = 0.4
young~age#58 = 0.1
young~age#34 = 0.8
young~age#20 = 1.0

suspect(id1,male,age#58,black,dark,height#180)
suspect(id2,female,age#34,green,dark,height#170)
suspect(id3,male,age#20,blue,brown,height#175)

?.-(X,_,young,light,_,tall).

X=id1 with 0.1 ; X=id2 with 0.4; X=id3 with 0.6
```

id2 and id3 are candidates to be guilty

Pattern matching recognition

- Recognize forge/fake copies from coins, bills, paintings, etc.

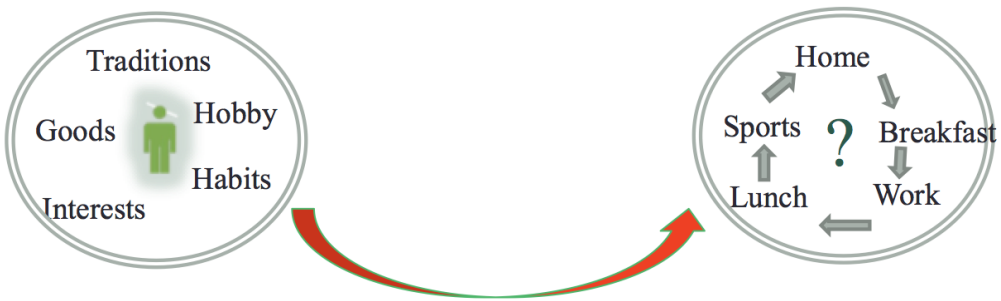
A lot of digital data

- Infrared scaner
- Ultraviolet scaner
- Ultrasound scaner
- Digital microscopes
- etc.

Apply Mathematics, AI and AR tools to the data in order to design intelligence systems to Digital Forensic

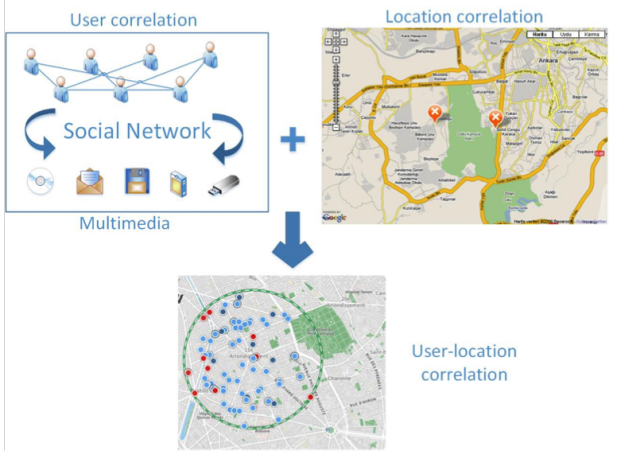
Recommendation systems

- Know **the best places in which we are interested in**, is a usual goal.
- A lot of information in internet mades **difficult this task**.
- **Recommendation systems** are very useful for both the users and the service providers. For example, for recommending movies, books, news, social networks, etc.



Social Network Analysis

- **Social Network Analysis** is one of these areas interested in developing recommendation systems, since they have access to a lot of information on their users.
- **Location-base social networks** (LBSN) are used to share information related to the activities one person is doing via geo-tagged user-generated multimedia content.
- This kind of social networks can store much information about the locations of their users, as well as information about such **places** and the **schedules** in which these spots are frequented.



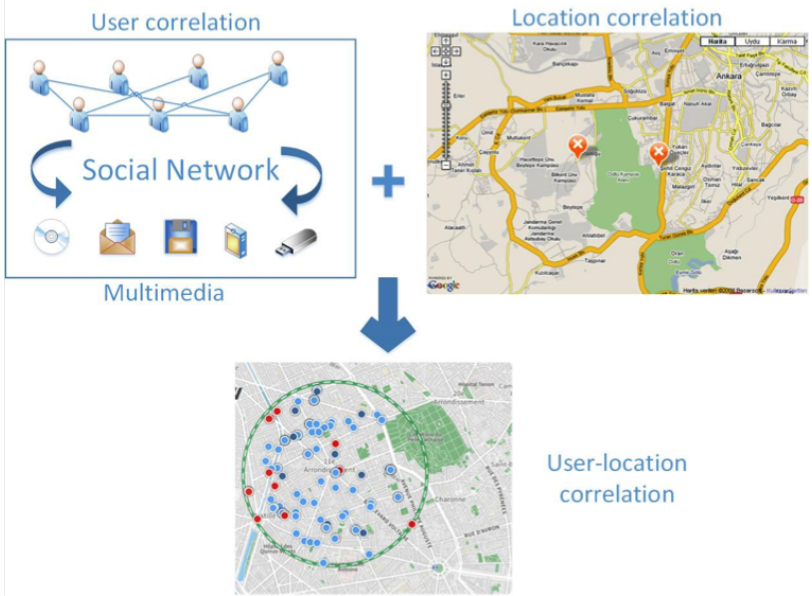
Formal Concept Analysis (FCA)

- FCA is a powerful tool to deal with several problems related to **data analysis**.
- Specifically, FCA analyzes data from a **relational database** that contain a set of objects and a set of attributes.
- Due to its **large number of applications**, FCA has gained a great popularity in the recent years and numerous generalizations of this theory have appeared.
- One of the most general and flexible frameworks is **multi-adjoint concept lattice**, in which a fuzzy environment is considered.
- We can handle **uncertainty, imprecise data or incomplete information**, which are important features in the recent years.
- A **wider range of applications** can be considered.

Locate-Based Social Network to DF

Definition

A LBSN (G, C) consists of a social network $G = (U, E)$, where U is the set of users, $E \subseteq U \times U$ is a relation between users. $C \subseteq U \times L \times T$ is a relation in which each element $(u, l, t) \in C$ represents the location $l \in L$ at time $t \in T$ of one user $u \in U$. Every element of C is called **check-in**.

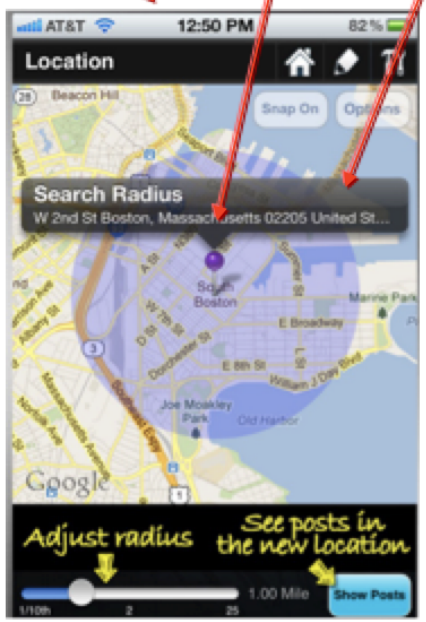


This information stored as check-in in the LBSN describes the **user's interest and habits**. Analyzing LBSN we can **discover behaviors and customs** about the users.

Foursquare data set

{User ID, Time, Location (x,y), Topic}

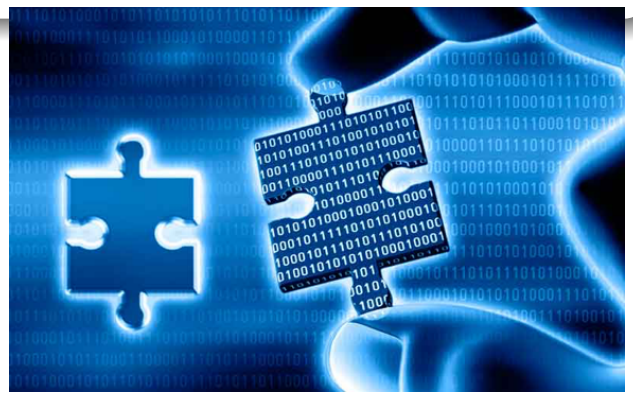
User ID	Data, Time	Loc Name	Loc(x,y)	Topic
1	Jul29 2011 17:38:54	Bobby Van's	40.7609966 -73.982902	Nightlife Spot,Food
1891	Jun27 2010 1:05:45	Union Hall	40.67625243 -73.9801526	Nightlife Spot,Music Venue
5	May12 2011 23:33:12	32nd. Korean Way	40.74822004 -73.9877807	Food
882	Mar13 2011 17:18:25	Koreatown	34.06173469 -118.300749	Outdoors & Recreation
882	Nov 14 2010 03:37:48	Fry's Electronics	34.19098603 -118.350169	Shop & Service
884	Feb 16 2011 04:18:41	Pedal Spin Studio	34.20496571 -118.226457	Service,Gym, Fitness Center



Fuzzy Logic and Digital Forensics

Currently, fuzzy logic is actively employed into the digital forensics area

- Intrusion detection, pass authentication, image forensics
- Network forensics based on fuzzy logic
- Automated generation of fuzzy rules from large-scale network
- Traffic analysis in digital forensics investigations
- Computer crime investigation by means of fuzzy semantics maps
- Anomaly detection using fuzzy association rules



Otrer U. of Cádiz competences

Legal boundaries and data protection

- Analysis of technology impact on laws
- Guidelines of data retention
- The evidence in Digital Forensics

- GRDP: the European visión
- Shock between protección and information flow

Ethical & Societal Impact

Impact assessment

- Societal impact
- Relationship with end users
- Efficiency & Efficacy index

Research Ethical Guidelines

- Methodological assessment

COST Action – DigForASP

DIGital FORensics: evidence Analysis via intelligent Systems and Practices - CA17124

- European Cooperation in Science & Technology (COST), funded by Horizonte 2020.
- From 10/09/2018 to 09/09/2022
- Action Chair: Jesús Medina
- 33 countries and more than 120 researchers.

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

DIGforASP

Funded by the Horizon 2020 Framework Programme of the European Union

Infoday – DigForASP

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

REGIONAL INFO DAY - KRAKOW 2019

Digital forensics: evidence analysis via intelligent systems and practices

CA17124

Tomorrow, from 10 to 13:30, and from 14:30 to 15:30

Thank you Prof. Piotr A. Kowalski

Infoday – DigForASP

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

REGIONAL INFO DAY - KRAKOW 2019

Digital forensics: evidence analysis via intelligent systems and practices

CA17124

Tomorrow, from 10 to 13:30, and from 14:30 to 15:30

Thank you Prof. Piotr A. Kowalski

Mathematical tools for Digital Forensics

Jesús Medina

Action Chair DigForASP – CA17124
Department of Mathematics. University of Cádiz, Spain
Email: jesus.medina@uca.es

UCA Universidad de Cádiz
Departamento de Matemáticas

Krakow, Poland
March 29th, 2019

DIGforASP