Introduction to Formal Concept Analysis

Bastian Wormuth / Peter Becker

2nd International Conference of Formal Concept Analysis
February 23 – February 27, 2004
Sydney, Australia
Outline for the Tutorial:

- Presentation
- Short Break
- Practical Session
Outline for the Presentation:

• What is FCA?
• Motivation
• What is a Formal Concept?
• Formal Contexts
• Data Sets
• The Concept Lattice: Reading, drawing, interpretation
• Conceptual Scaling
• Summary
Formal Concept Analysis is ... 

... is a mathematization of the philosophical understanding of concept

... a human-centered method to structure and analyze data

... a method to visualize data and its inherent structures, implications and dependencies
Motivation

... example from ToscanaJ
What is a concept?

Let's examine an example, the concept “car”:

What drives us to call an object a “car”?

• Every object having certain attributes is called “car”:
  • a car has tires
  • a car has a motor
  • a car has the purpose of transportation
  • a car has seats ... etc.

• All objects having these attributes are called “cars”:
  • Mercedes, Nissans, Toyotas are cars
  • San Francisco's Cable Car is called car ...
What is a concept?

This description of the concept “car” is based on sets of objects related to attributes:

Objects

Mercedes
Cable Car
Nissan
Toyota
...

Attributes

has motor
has tires
has seats
...

Objects, attributes and a relation form a concept.
What is a concept?

So the concept is constituted by two parts ...

... having a certain relation:

- all objects belonging to this concept have all the attributes of \( B \)
- all attributes belonging to this concept are shared by all objects of \( A \)
- \( A \) is called the concept's extension, \( B \) is called the concept's intension
The sets of a concept are embedded in their context:

These sets and their relations are the basis for all conclusions we make – any concept we derive, any implication we deduce is based on the context - it is our universe of discourse.

-> Changing the context will change the concepts and their structure
The Formal Context – Universe of Discourse

\( M \): a set of attributes

\( G \): a set of objects

<table>
<thead>
<tr>
<th></th>
<th>small</th>
<th>medium</th>
<th>big</th>
<th>twolegs</th>
<th>fourlegs</th>
<th>feathers</th>
<th>hair</th>
<th>fly</th>
<th>hunt</th>
<th>run</th>
<th>swim</th>
<th>mane</th>
<th>hooves</th>
</tr>
</thead>
<tbody>
<tr>
<td>dove</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>hen</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>duck</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>goose</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>owl</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>hawk</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>eagle</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>fox</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>dog</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>wolf</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>cat</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>tiger</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>lion</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>horse</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>zebra</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>cow</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>x</td>
</tr>
</tbody>
</table>

An incidence relation \( I \) between \( G \) and \( M \)

Such a group \((G, M, I)\) of objects \( G \), attributes \( M \) and a relation \( I \) is called a formal context.

There are many ways to transform more complex data into formal contexts.
The Formal Context – Universe of Discourse

**M**: a set of attributes

<table>
<thead>
<tr>
<th></th>
<th>small</th>
<th>medium</th>
<th>big</th>
<th>twolegs</th>
<th>fourlegs</th>
<th>feathers</th>
<th>hair</th>
<th>fly</th>
<th>hunt</th>
<th>run</th>
<th>swim</th>
<th>mane</th>
<th>hooves</th>
</tr>
</thead>
<tbody>
<tr>
<td>dove</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hen</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>duck</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>goose</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>owl</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td></td>
<td>.</td>
<td></td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hawk</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eagle</td>
<td></td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fox</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dog</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td></td>
<td>x</td>
<td></td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wolf</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cat</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tiger</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lion</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>horse</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>zebra</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cow</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td></td>
<td>.</td>
<td></td>
<td>.</td>
<td>.</td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

**G**: a set of objects

Transposing the matrix, changing objects and attributes, creates the dual structure – the same diagram, but flipped top down.
The resulting implications and relations are the same.

Bastian Wormuth / Peter Becker
Data Sets: Definition of Formal Concepts

• For the mathematical definition of *formal concepts* we introduce the derivation operators “'”. For a set of objects $A$, $A'$ is defined as:
  $A' = \{\text{all attributes in } M \text{ shared by the objects of } A\}$
For a set of attributes $B$, $B'$ is defined as:
  $B' = \{\text{all objects in } G \text{ that have all attributes of } B\}$.

• We are looking for pairs of sets $(A,B)$ of objects and attributes that fulfill the conditions $A' = B$ and $B' = A$ and we call these pairs *formal concepts*.
Using the derivation operators we can derive *formal concepts* from our *formal context* with the following routine:

1) Pick a set of objects $A$.
2) Derive the attributes $A'$.
3) Derive $(A')'$.
4) $(A'',A')$ is a *formal concept*.

A dual approach can be taken starting with an attribute.
1) Pick any set of objects A, e.g. A={duck}.
2) Derive the attributes A'={small, two legs, feathers, fly, swim}
3) Derive (A')'={small, two legs, feathers, fly, swim}'={duck, goose}
4) (A'',A')=(\{duck, goose\},\{small, two legs, feathers, fly, swim\}) is a formal concept.
The Concept Lattice

• The formal concept \((A'', A') = (\{duck, goose\}, \{small, two legs, feathers, fly, swim\})\) is represented in a line diagram as a node:

• Consider another formal concept:
  \((B'', B') = (\{duck, goose, dove, owl, hawk\}, \{small, two legs, feathers, fly\})\).

• The formal concept \((A'', A')\) is called subconcept of \((B'', B')\) and \((B'', B')\) is called superconcept of \((A'', A')\).

• \((A'', A')\) is drawn below \((B'', B')\) and connected with a line.
The Concept Lattice

- By adding more *formal concepts* the diagram is extended step by step:
  - ({owl, hawk}, {feathers, two legs, small, fly, hunt})
  - ({owl, hawk, eagle}, {feathers, two legs, fly, hunt})
  - ... plus the relations ...

- ... and so on ...

- Several methods exist to derive all *formal concepts*:
  - Ganter's algorithm, cut over extents, Lindig's algorithm etc.
The Concept Lattice

The subconcept – superconcept relation defines an order on the set $B$ of all formal concepts of a formal context.

- For two concepts (A,B) and (C,D) this order is formalized as:
  
  $$(A,B) \preceq (C,D) : \iff A \subseteq C \land D \subseteq B$$

- (A,B) is smaller than (C,D) if A is subset of C (and D is subset of B)

- $(B,\leq)$ is an ordered set.

The set $B$ of formal concepts has another property:

- for each set of formal concepts of a formal context there exists always a unique greatest subconcept (meet) and a unique smallest superconcept (join).

The ordered set $(B,\leq)$ plus the last property forms a mathematical structure: the concept lattice.
The Concept Lattice

more specific concepts

more general concepts
The Concept Lattice

Labelling conventions:

- The greatest concept having a certain attribute \(a\) in its intent gets the label \(a\) (attribute concept).
- The smallest concept having a certain object \(o\) in its extent gets the label \(o\) (object concept).

This allows the reconstruction of the whole extents and intents for every concept.
The Concept Lattice: Conjunction of concepts VI

Looking for objects with certain attributes:

Which object has the attributes \(a_2\) and \(a_3\)?

Find the attribute concepts having \(a_2\) and \(a_3\) as intent and follow the lines down to the concept where they meet.
The Concept Lattice: Disjunction of concepts VII

Looking for attributes that have certain objects in common:

Which attributes are shared by objects $o_8$ and $o_6$?

Find the concepts having $o_8$ and $o_6$ as extent and follow the lines up to the concept where they join.
The *Concept Lattice*: Implications

Implications could also be identified:

Every object that has attribute \( a_5 \) also has attribute \( a_3 \). So \( a_5 \) implies \( a_3 \).
The Concept Lattice: Implications

Waters and their attributes:

Running implies inland and constant
Conceptual Scaling

• A diagram based on a subset of attributes of a formal context is called a conceptual scale.
• The process of creating single-valued contexts from a many-valued data set is called conceptual scaling.
• Conceptual scaling could be standardized – but relies mostly on the human interpretation.
• The diagrams layout also influences the interpretation, different layouts might emphasize different aspects.
• Typically there is not a single best layout. It often depends on purpose and reader.
• There are some drawing “rules” to keep the layout neutral.
• Drawing a nice lattice automatically is an unsolved problem and still an area of research.

• Next: some standard scales …
Conceptual Scaling: Nominal Scale
Conceptual Scaling: Ordinal Scale

III

Bastian Wormuth / Peter Becker
Conceptual Scaling: Biordinal Scale

Bastian Wormuth / Peter Becker
Conceptual Scaling: Interordinal Scale

![Diagram of a concept lattice with various thresholds and nodes labeled with numbers.]
Conceptual Scaling: Boolean Scale
Conceptual Scaling: Boolean Scale
Conceptual Scaling: Boolean Scale
Annotation:

The scales shown represent typical data structures
-> you will recognize them or parts of them in your data
Conceptual Scaling: Annotation

Waters and their attributes:
Lattice Drawing: Approaches

- Avoid crossing of lines.
- Try to draw parallel lines.
- Identify known structures: cubes, rectangles ...
- Layer the nodes: draw nodes on the same layer, if their concept's extents have the same size.
- Try to draw steep lines, avoid flat ones.
Lattice Drawing: Additive Line Diagrams II

- One approach is to assign vectors to the attributes.
- Each node's position is determined by the sum of all vectors of attributes in their concept's intent.
Summary

- Exploring the data: „playing“ with the data
- Regarding everything embedded in its conceptual surrounding
- The diagrams are stimulating discussions
- Identifying knowledge gaps and mistakes in the database
- FCA is creating a picture of people’s imagination
- ... could support data guided theory building
References / Sources:

Literature:
• B. Ganter/R. Wille: Formal Concept Analysis, Mathematical Foundations
• R. Wille: Introduction to Formal Concept Analysis

WWW:
• http://www.kvocentral.org
• This presentation could be downloaded via the conference homepage or at http://www.wormuth.info
• http://www.tockit.org

Email:
• bastian@wormuth.info
• peter@peterbecker.de
Thank you!