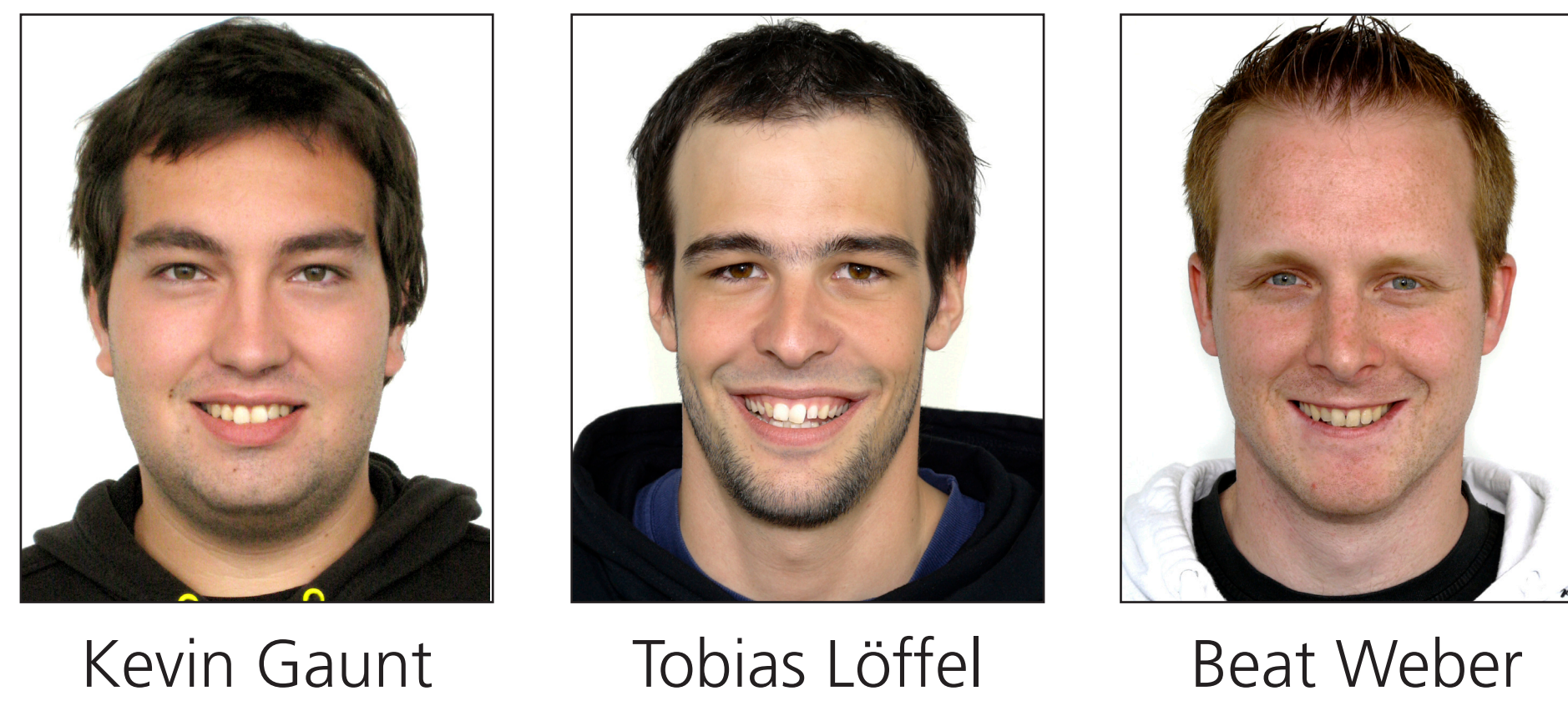


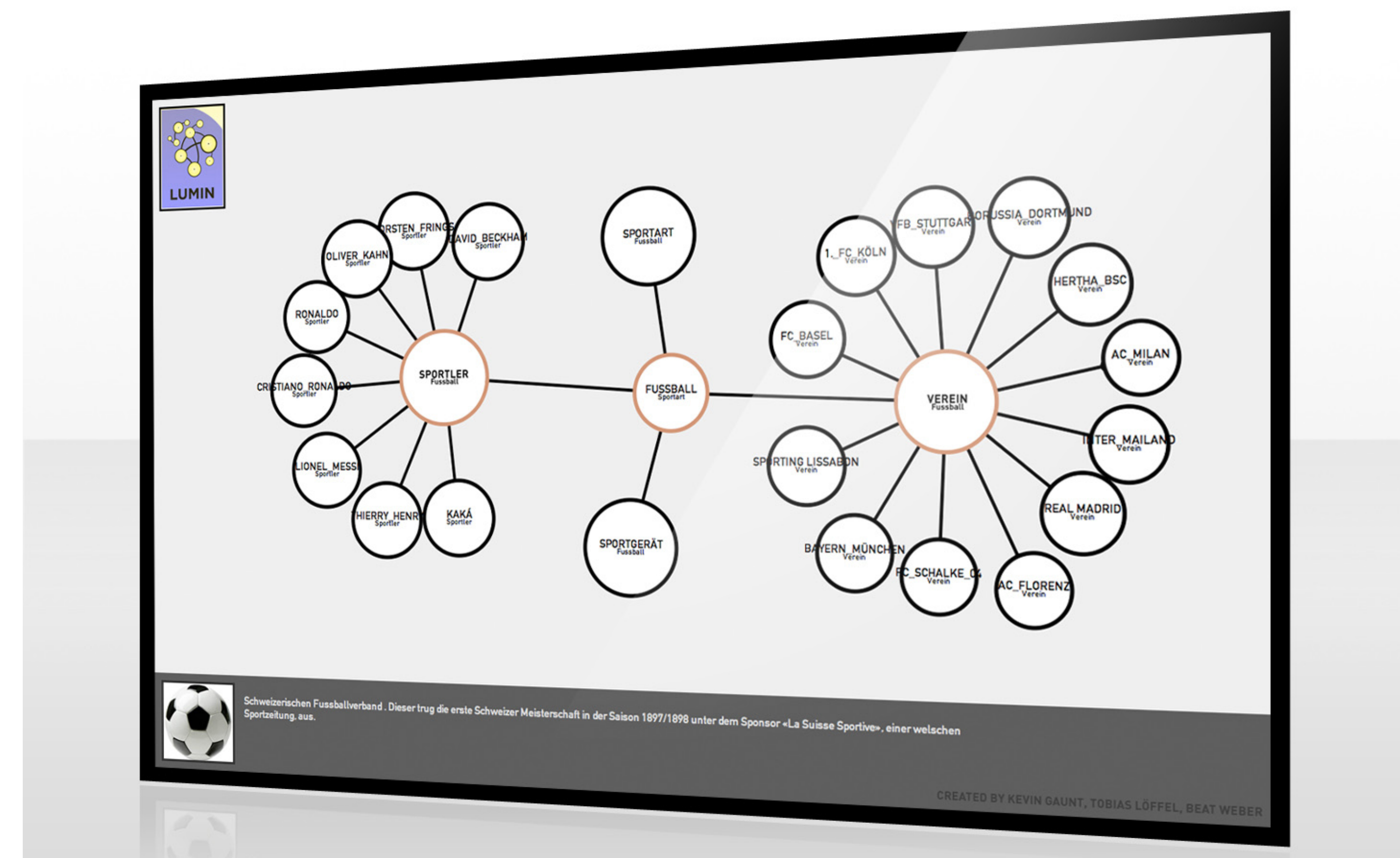
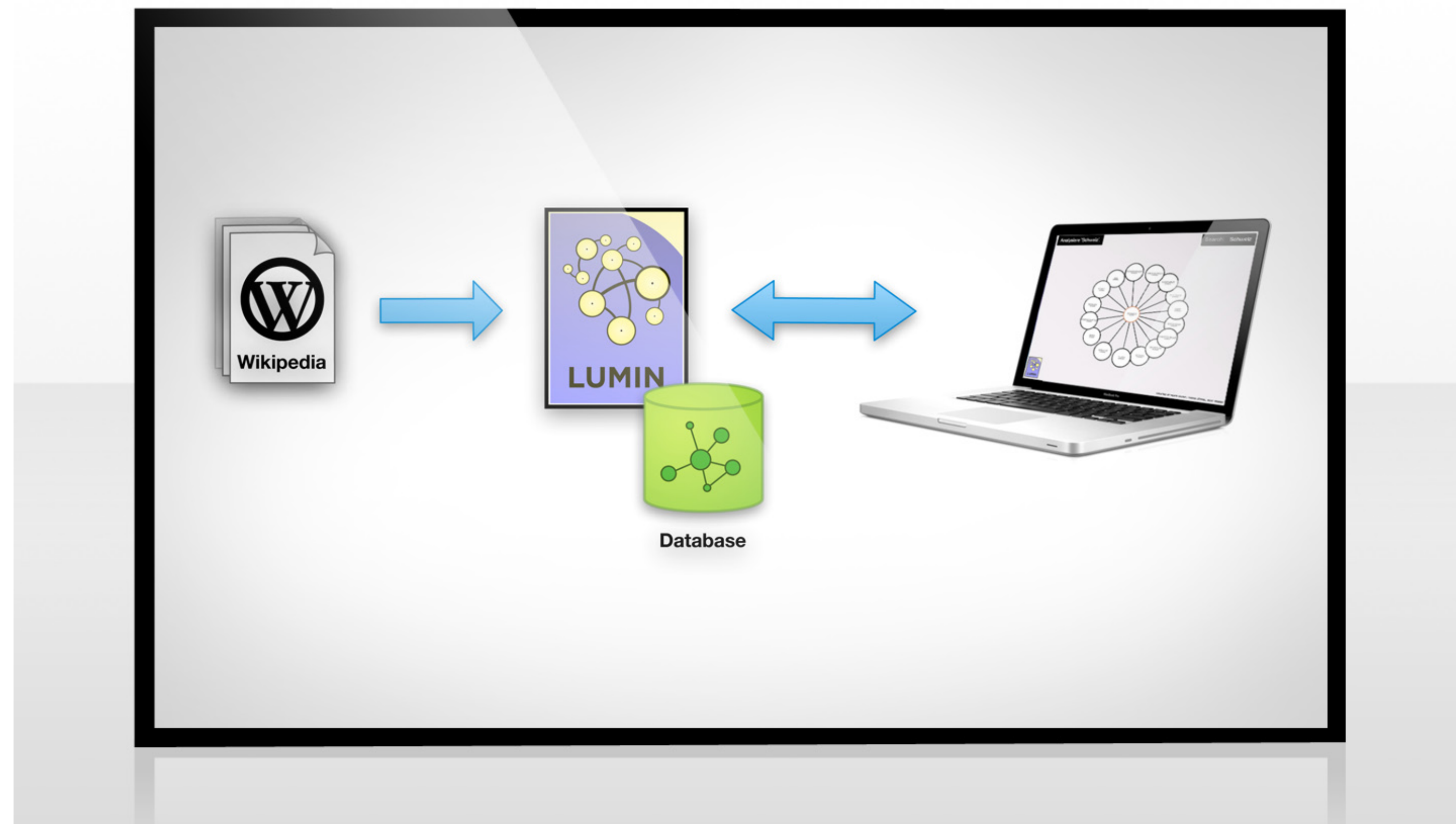
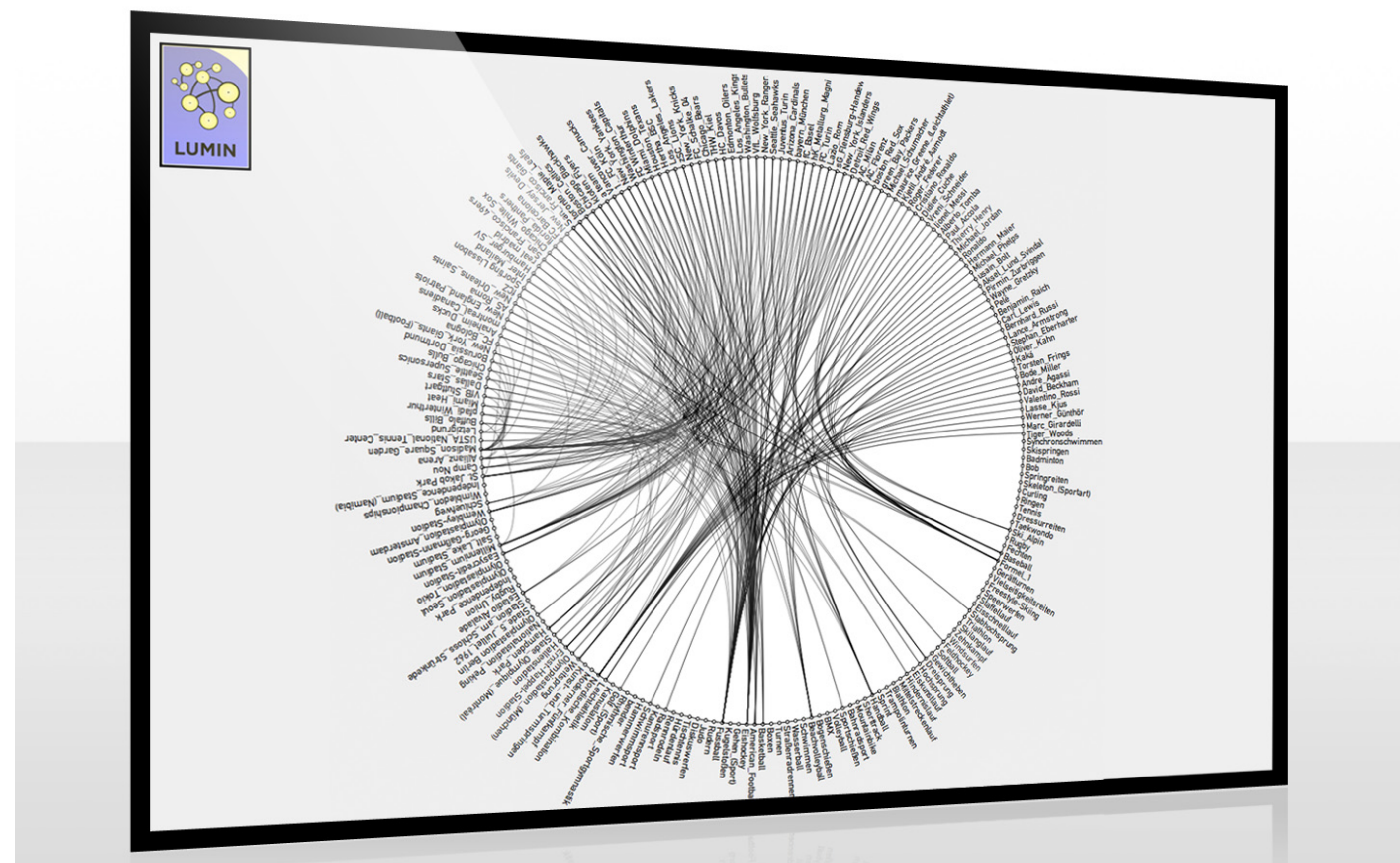
PROJEKT LUMIN

Wissensextraktion aus im Internet verfügbaren Informationsquellen



Betreuer

Prof. Dr. Josef Joller



Ausgangslage

Während das Internet für uns Menschen zu einer unabdingbar wichtigen Quelle für Informationen aller Art wurde, kann der Computer nichts mit dem, auf mittlerweile fast 29 Milliarden Webseiten publizierten, Wissen anfangen. Dieses für den Computer erfassbar zu machen, bedingt die darin enthaltenen Informationen in eine abstraktere Struktur umzuwandeln. Die grosse Vision dieser Bachelorarbeit also ist es, einen Prototypen zu entwickeln, welcher das vorhandene Wissen soweit gliedert, dass es maschinen-verstehbar wird.

Resultate

Die Bachelorarbeit befasst sich unter anderem mit den Themengebieten 'Knowledge Extraction', 'Text Mining' und 'Clustering'. Es wurde ein System entwickelt, welches mit einer Erfolgsquote von über 95% (bei rund 10% vorgegebenen Daten), kontextuell verwandte Begriffe einander zuordnen und deren Typ bestimmen kann. Sämtliche Terme werden in der graph-basierten, skalierbaren Datenbank Neo4j persistiert, welche besonders auf Lese- und Traversierungs-Operationen optimiert ist. Ein optisch ansprechendes Graphical User Interface

präsentiert die gewonnen Informationen auf verständliche Art und Weise. Durch den Einsatz ausschliesslich plattformunabhängiger Technologien, ist das System auf allen - von Java 6 unterstützten Betriebssystemen - einsatzfähig, sogar in heterogenen Konfigurationen.

Mögliche Anwendungen

Mögliche Anwendungen liegen in der Erstellung alternativer Suchmaschinen, die auf den erstellten semantischen Informationen operieren. So wäre es beispielsweise möglich, alle Städte der Schweiz anzuzeigen, welche einen Löwen im Wappen haben, ohne dass dafür eine spezifische Informationsseite indexiert werden muss. Ausserdem könnte eine weiterentwickelte Version den Menschen bei der Ausführung einfacher Recherche-Tasks unterstützen.

Technologien

Eine Vielfalt verschiedener Technologien wurde in diesem Projekt eingesetzt. Eine Auswahl wird in der untenstehenden Liste kurz aufgeführt.

- Java 6
- Neo4j
- Processing 1.0
- JMS
- XML

Weitere Informationen

Gerne beantwortet das Projektteam Ihre Fragen oder gibt Ihnen eine kurze Präsentation. Bei Interesse melden Sie sich mit Vorteil per E-Mail.

- Kevin Gaunt, kgaunt@lumin.ch
- Tobias Löffel, tloeffel@lumin.ch
- Beat Weber, bweber@lumin.ch

1

Suchbegriff eingeben

2

Informationen beschaffen

3

Wissen extrahieren

4

Wissen analysieren

5

Begriff clustern

6

Resultate visualisieren