Learning Analytics in Portalen für Fachinformationen

Prof. Dr. Hendrik Drachsler
@hdrachsler
WhoAmI

- Hendrik Drachsler
  Professor Educational Technologies & Learning Analytics

- Research topics:
  Recommender Systems
  Learning Analytics

- Application domains:
  Schools
  HEI
  Medical education
1. Definition of trust and Learning Analytics
2. Fears of Learning Analytics
3. Humboldt-ian Educational Model
4. New demands from GDPR 2018
5. Technical approaches towards Trusted Learning Analytics
6. Learning Analytics in Info portals
A definition of Trust

How do you define trust?

Picture by Terry Johnston
https://www.flickr.com/photos/powerbooktrance/466709245/
A definition of Trust

Trust is about a firm belief in the reliability, truth, or ability of someone or something.

A trustful relation is mutually based on
• openness
• truth
• reliability
• integrity
• belief
• faith
• freedom of suspicion

Picture by Terry Johnston
https://www.flickr.com/photos/powerbooktrance/466709245/
What are Learning Analytics for you?
A definition of Learning Analytics

TRUSTED LEARNING ANALYTICS
New insights

Graph by Rob Koper. Data science voor de realisatie van online activerend onderwijs. Presentation given at Dag van het Onderwijs (5 November 2015). Heerlen. The Netherlands
New insights

Graph by Rob Koper. **Data science voor de realisatie van online activerend onderwijs.**
Presentation given at Dag van het Onderwijs (5 November 2015). Heerlen. The Netherlands
Lecture structure

1. Definitions
2. Fears of Learning Analytics
3. Humboldt-ian Educational Model
4. New demands from GDPR 2018
5. Technical approaches towards Trusted Learning Analytics
6. Learning Analytics in Info portals
Do you have any concerns when you think about Learning Analytics in K12 or HEI?
People are afraid of AI and digital technologies
Taxi and cab drivers and chauffeurs

Likelihood of automation?
It's too close to call (57%)

How this compares with other jobs:
159th of 366

0% 25% 50% 75% 100%
Most likely Least likely

Share my result
Learning Analytics has a trust problem ...
... because Learning Analytics has the potential of becoming a high stakes assessment.
Black box vs. White box

Unknown algorithms
Unknown data collection
Automated decisions
No access to raw data
No control who uses it

Open algorithms
Transparent indicators
No automated decisions
Full access to data
Knowing who accesses your data
OpenSCHUFA – shedding light on Germany’s opaque credit scoring

Why we started OpenSCHUFA, why you should care about credit scoring & how you can help.

AlgorithmWatch on February 21st, 2018
Educational Example

Ignoring the fears and public perception of the application of analytics can lead to a lack of acceptance, protests, and even failure of entire LA implementations.
Is the German Education system prepared for the challenges of the data-driven society?
Humboldt-ian Model in the age of Big Data

Designed an education system …
- based on unbiased knowledge (objective analytics but also critical reflection)
- combining research and teaching
- allowing students to choose their own course of study (personalization)
- develop autonomous reflected individuals (self-regulated learning)
- education beyond vocational training only
- comprehensive general learning (lifelong learning)
- cultural knowledge (formative feedback vs. summative assessment)
Digitaler Humanismus

Julian Nida-Rümelin, Nathalie Weidenfeld

Eine Ethik für das Zeitalter der künstlichen Intelligenz

Sachbuch

PIPER
Lecture structure

1. Definitions
2. Fears of Learning Analytics
3. Humbold-tian Educational Model
4. New demands from GDPR 2018
5. Technical approaches towards Trusted Learning Analytics
6. Learning Analytics in Info portals
The EU General Data Protection Regulation (GDPR) is the most important change in data privacy regulation in 20 years - we're here to make sure you're prepared.

“the biggest change to data protection law for a generation”
Elizabeth Denham, Information Commissioner
GDPR 2018

- Right to be informed
- Right of access
- Right to rectification
- Right to erasure
- Right to restrict processing
- Right to data portability
- Right to object automated decision making

Do your Learning Technology systems support these rights?
Some things are already on its way

<table>
<thead>
<tr>
<th>D</th>
<th>DETERMINATION — Why you want to apply Learning Analytics?</th>
<th>What is the added value (organisational and data subjects)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>EXPLAIN — be open about your intentions and objectives</td>
<td>What data will be collected for which purpose?</td>
</tr>
<tr>
<td>L</td>
<td>LEGITIMATE — Why are you allowed to have the data?</td>
<td>Which data sources have already (aren’t they enough)?</td>
</tr>
<tr>
<td>I</td>
<td>INVOLVE — Involve all stakeholders and the data subjects</td>
<td>Provide access to the personal data collected (about the data subjects)?</td>
</tr>
<tr>
<td>C</td>
<td>CONSENT — Make a contract with the data subjects</td>
<td>Ask for a consent from the data subjects before the data collection</td>
</tr>
<tr>
<td>A</td>
<td>ANONYMISE — Make the individuals not retrievable</td>
<td>Anonymise the data as far as possible</td>
</tr>
<tr>
<td>T</td>
<td>TECHNICAL — Procedures to guarantee privacy</td>
<td>Aggregate data to generate abstract metadata models (those do not fall under EU Directive 95/46/EC)</td>
</tr>
<tr>
<td>E</td>
<td>EXTERNAL — If you work with external providers</td>
<td>Make sure the data storage fulfills international security standards</td>
</tr>
</tbody>
</table>


Online at: http://www.laceproject.eu/ethics-privacy/
Lecture structure

1. Definitions
2. Fears of Learning Analytics
3. Humboldt-tian Educational Model
4. New demands from GDPR 2018
5. Technical approaches towards Trusted Learning Analytics
6. Learning Analytics in Info portals
Trusted Learning Analytics Infrastructure

- T-4-LA first GDPR 2018 conform Big Data Infrastructure
- Joined project with GU und OU
- Among ‘traditional’ learning data we also aim to collect multimodal data.

EC-TEL conference 2018, Leeds, UK.
Learning Analytics Indicator Repository

Lecture structure

1. Definitions
2. Fears of Learning Analytics
3. Humbold-tian Educational Model
4. New demands from GDPR 2018
5. Technical approaches towards Trusted Learning Analytics
6. Learning Analytics in Info portals
Learning Analytics in information portals

Educational Fingerprinting
Information portals and learning

- Information portals offer a broad variety of information
- Users interact with content to gain information / knowledge
How can we faster adapt information portals to individual learners?

• **Goal:** User centric adaptive (open) educational resources
• **Problem:** First time users without user account
• **Needed:** Previous Knowledge – without manual user input
Learner Modeling

- Learner Modeling = Data Collection + Profile Construction (Brusilovsky)

- Current approaches:
  - Learner accounts
  - Ethernet or browser proxy
  - Browser plugin
Current Data Collection Problems

• Interpretation requires data of several sessions
• Accounts require learner login
• Learning Analysis Proxies or Browser Plugins must be set up manually

To experience adaptive content, learners have to be …

1. active
2. continuous users
3. aware, that they can receive adaptive resources in the future
Fingerprinting

- Known web analytics technology for user identification
- Available ubiquitously
- Uses information of user configuration sent by the browser
Scientific Fingerprinting Projects

- Learners are unique due to their web fingerprint

<table>
<thead>
<tr>
<th>Observed Uniqueness</th>
<th>Panopticlick</th>
<th>Am I Unique?</th>
<th>Cao et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>470,161</td>
<td>118,934</td>
<td>1,903</td>
</tr>
<tr>
<td>Unique users</td>
<td>83.6%</td>
<td>89.0%</td>
<td>99.2%</td>
</tr>
<tr>
<td>Pub. year</td>
<td>2010</td>
<td>2016</td>
<td>2017</td>
</tr>
</tbody>
</table>

https://amiunique.org/

https://panopticlick.eff.org/
Example Fingerprint

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Similarity ratio</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>User agent</td>
<td>&lt;0.1%</td>
<td>&quot;Mozilla/5.0 (Windows NT 6.1; WOW64; /60.0'</td>
</tr>
<tr>
<td>Accept</td>
<td>51.12%</td>
<td>&quot;text/html,application/xhtml+xml,application/xml;q=0.9,<em>/</em>;q=0.8</td>
</tr>
<tr>
<td>Content encoding</td>
<td>46.22%</td>
<td>&quot;gzip, deflate, br&quot;</td>
</tr>
<tr>
<td>Content language</td>
<td>3.91%</td>
<td>&quot;de,en-US;q=0.7,en;q=0.3&quot;</td>
</tr>
<tr>
<td>List of plugins</td>
<td>&lt;0.1%</td>
<td>&quot;Plugin 0: Shockwave Flash; Shockwave Flash 15.0.0.203.480.0.dll.</td>
</tr>
</tbody>
</table>
Experimental Design

• **Research Question / Hypothesis:**
  - Can we provide personalised information to visitors of information portals based on their FP?
  - Information portal visitors can be modeled according to their FP.

• **Approach:**
  - Identify individual learning models by a questionnaire
  - Assign Fingerprinting data to learner models
  - Validate extracted correlations by a second user group

• **Possible Outcome:**
  - Different learning behaviour of users with advanced PCs vs. Users with outdated PCs.
Outlook / Vision

• Create a data driven „Bildungsindex“ fed by the search terms and FPs of different users that visit an information portal for education.
• Continuous expert survey to generate an national „Ifo Geschäftsklima Index“ for education

https://trends.google.de/
Many thanks for your attention!

Questions now or later:

@hdrachsler  drachsler@dipf.de