# Nebenbedingungen in der Personensuche im Internet

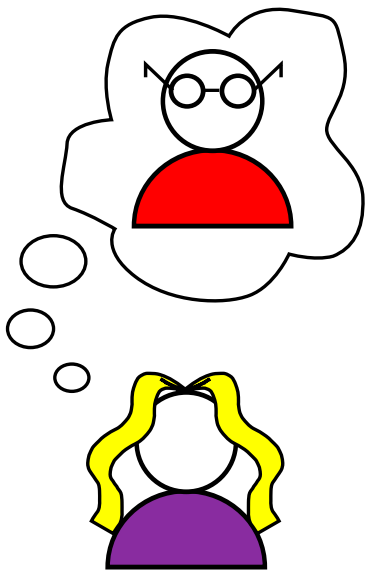
## Constraints in Web People Search

Johannes Kiesel

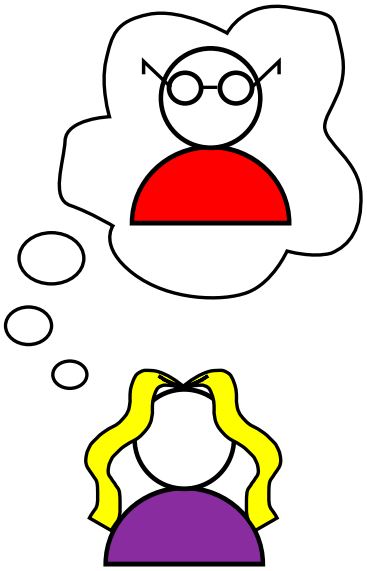
Bauhaus-Universität Weimar

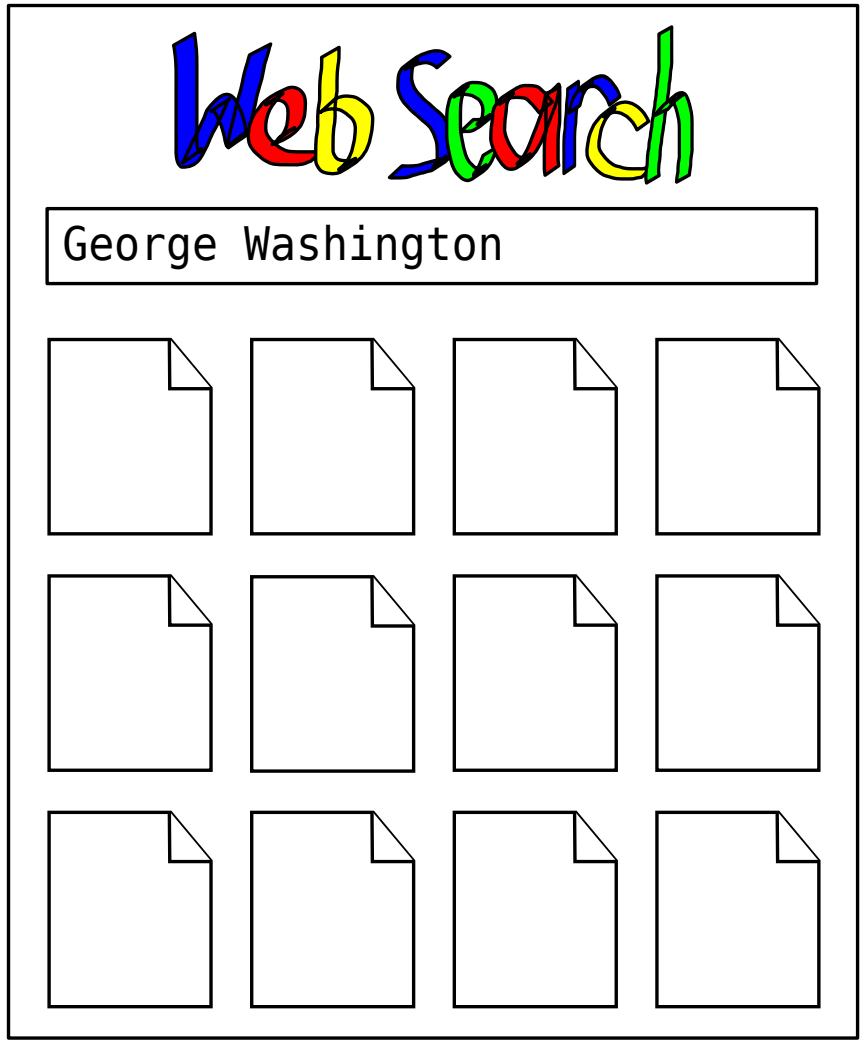
3. Oktober 2012

# Was ist „Personensuche im Internet“?

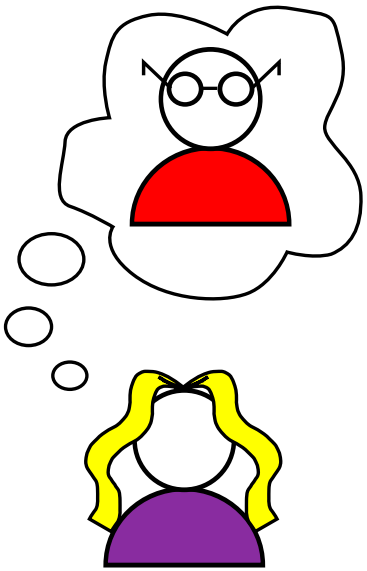


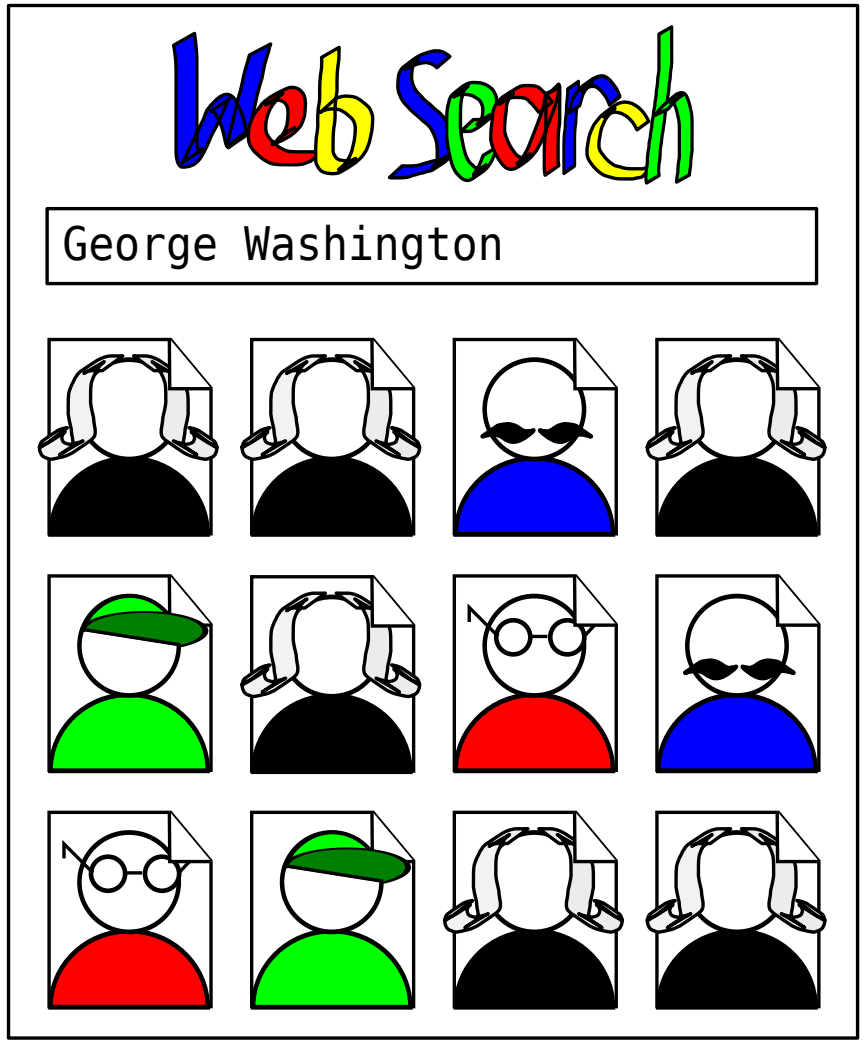
# Was ist „Personensuche im Internet“?



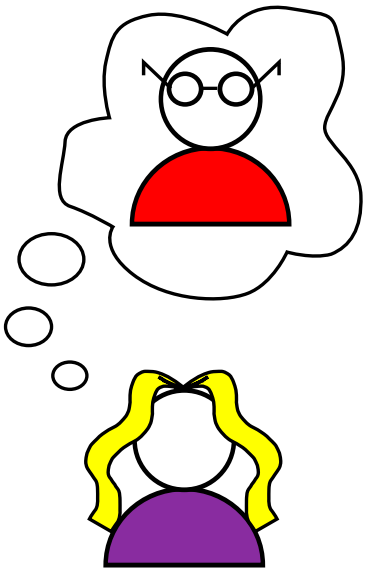


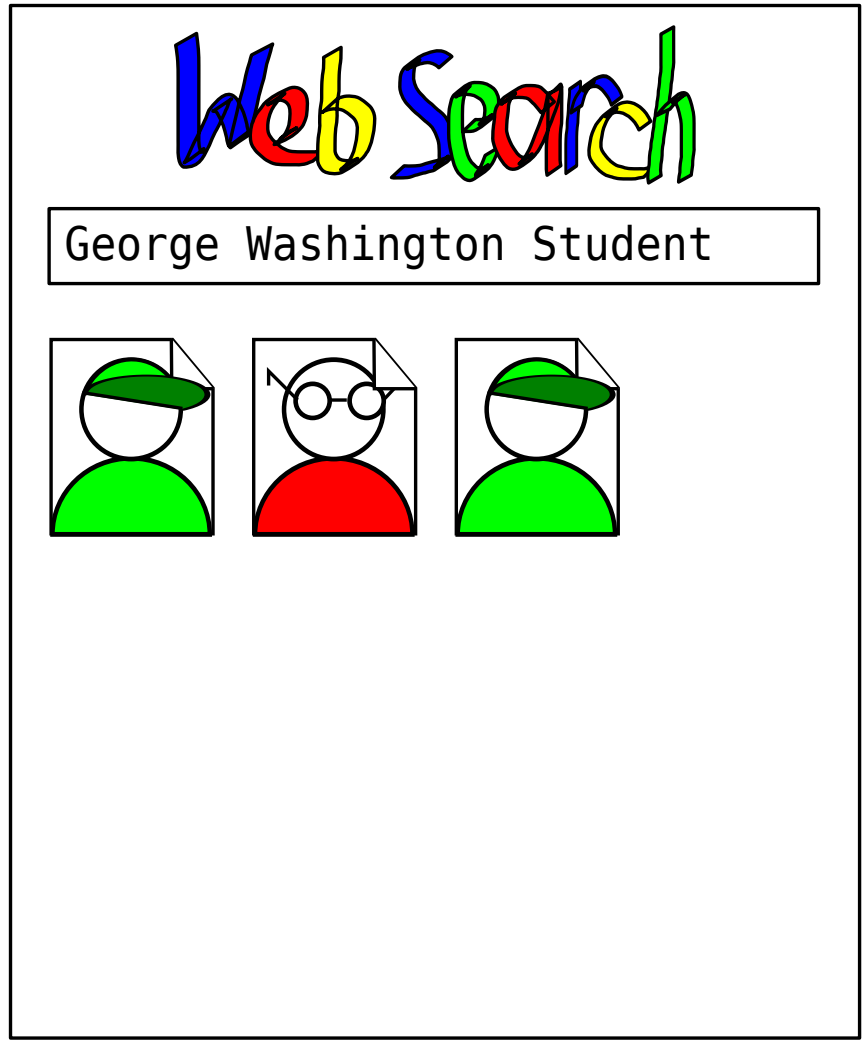
# Was ist „Personensuche im Internet“?





# Was ist „Personensuche im Internet“?

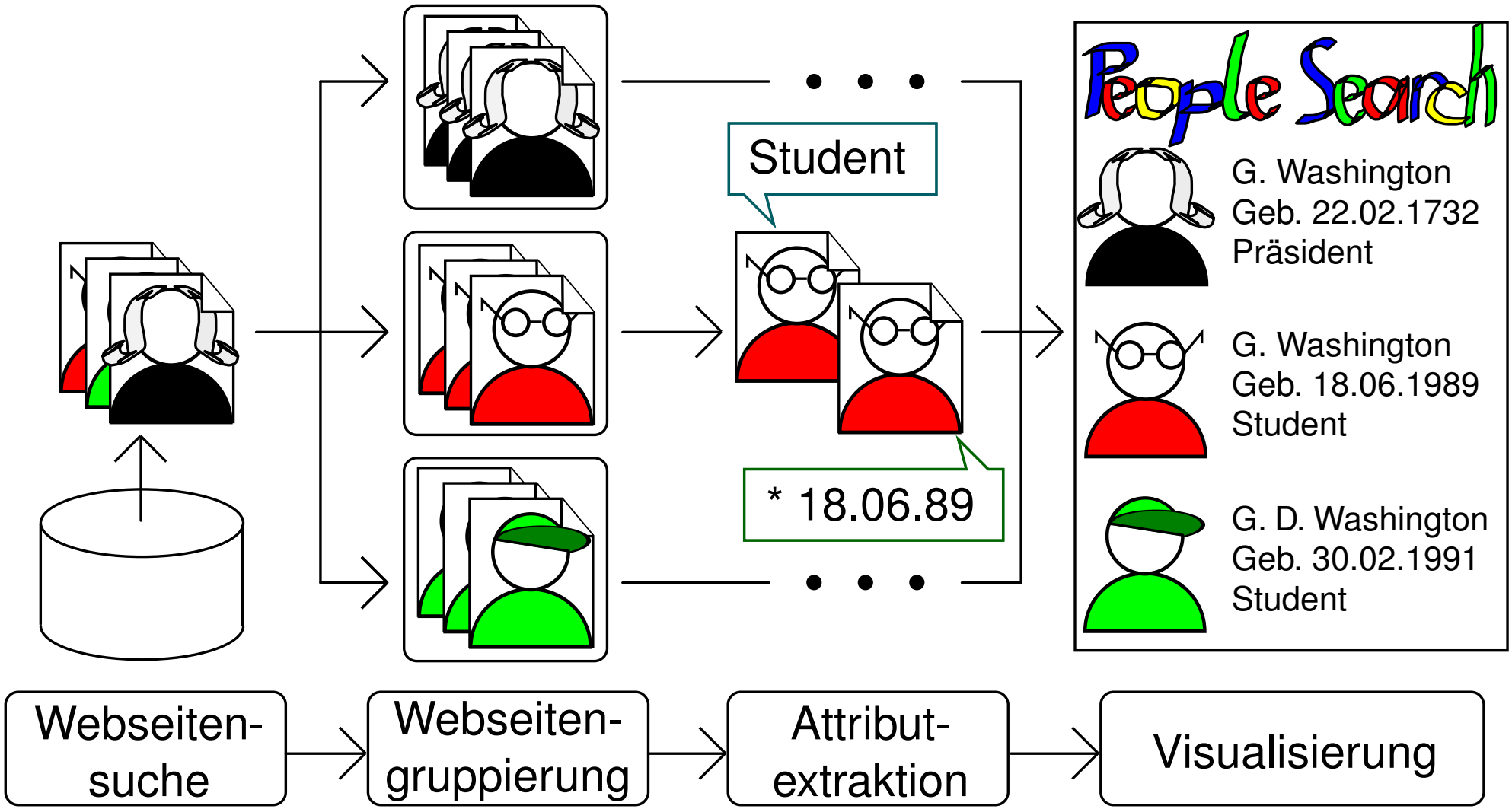




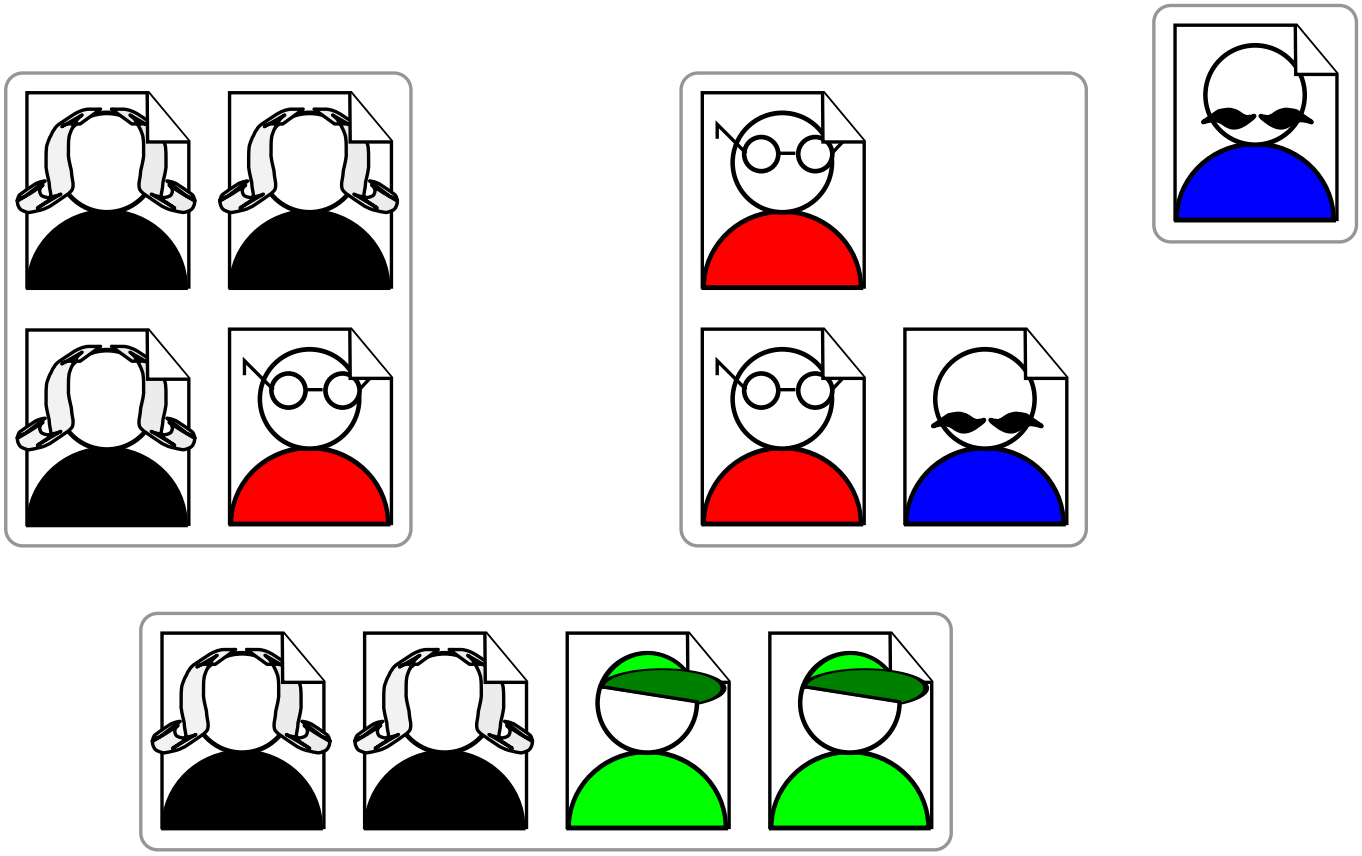
# Aktuelle Systeme


[http://www.zoominfo.com (Nachbearbeitet; Aufgerufen: 20.04.2012)]

# Vorgehensweise

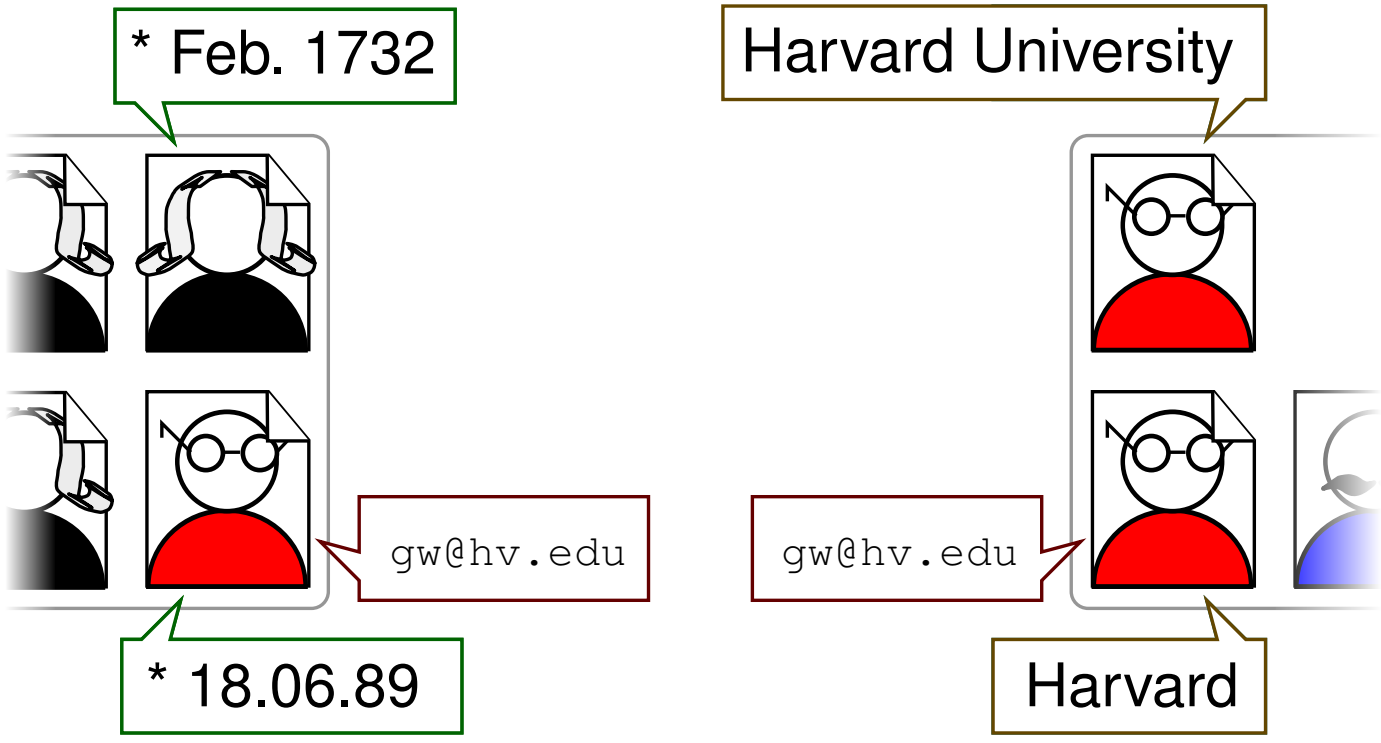


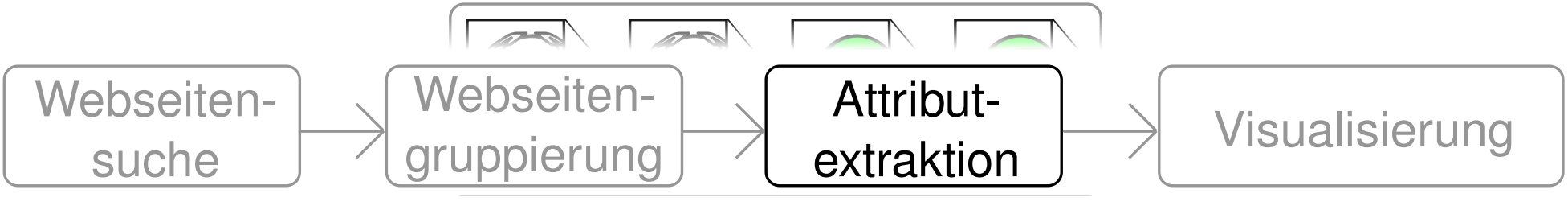
# Vorgehensweise



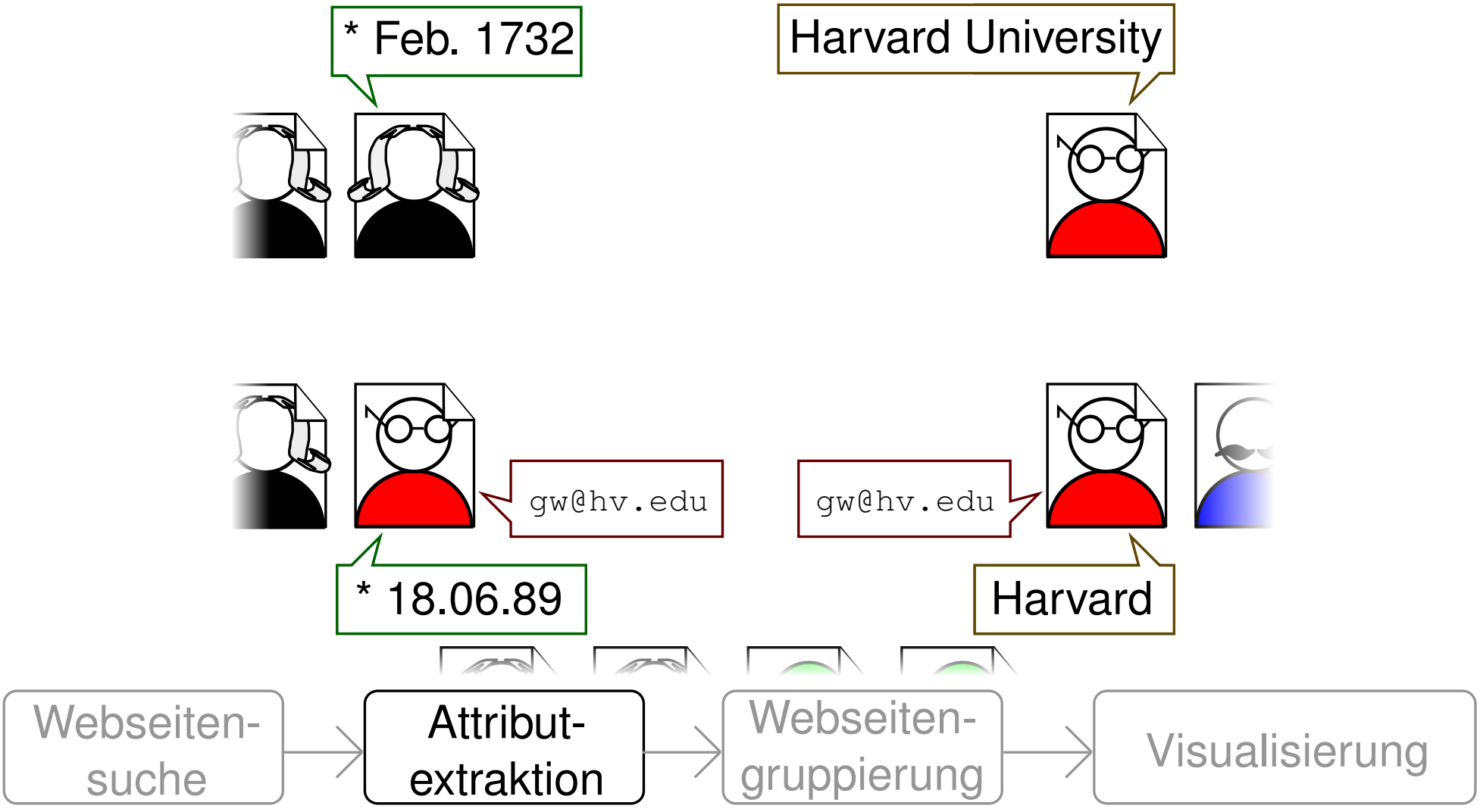
Webseiten-suche → **Webseiten-gruppierung** → Attribut-extraktion → Visualisierung

# Vorgehensweise



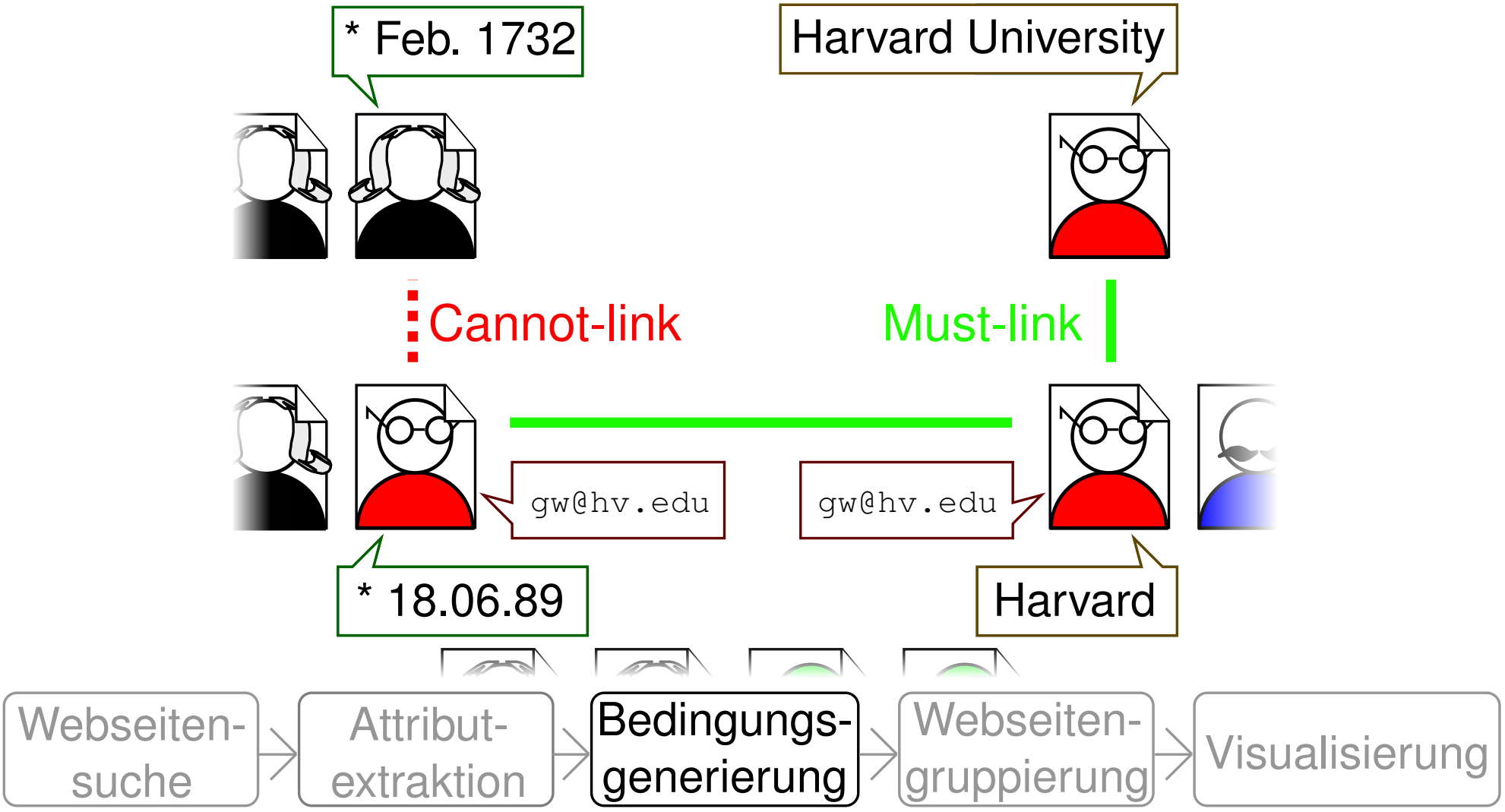



# Generierung von Nebenbedingungen



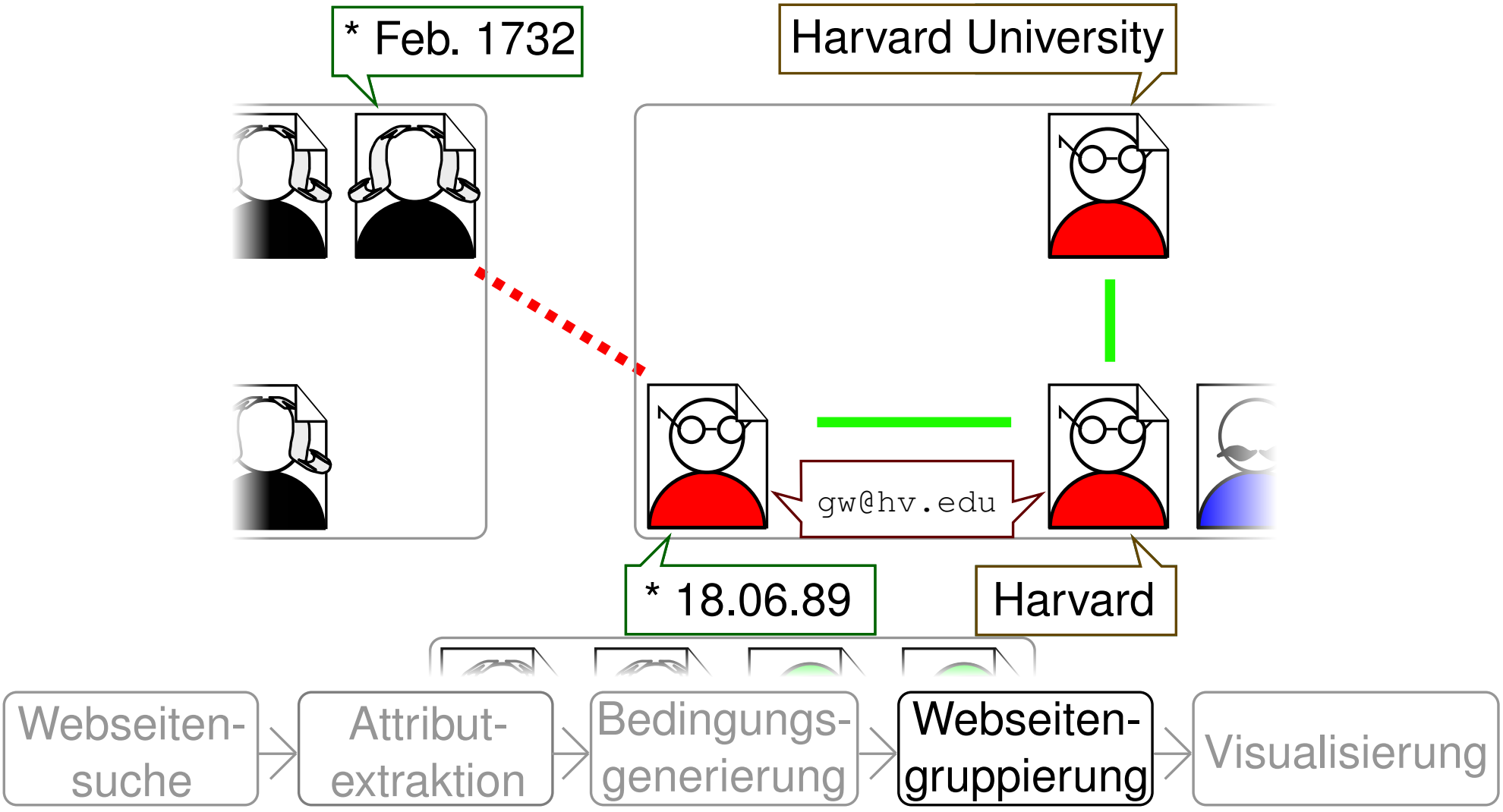

[K. Wagstaff, C. Cardie. Clustering with Instance-Level Constraints. 2000]

# Generierung von Nebenbedingungen



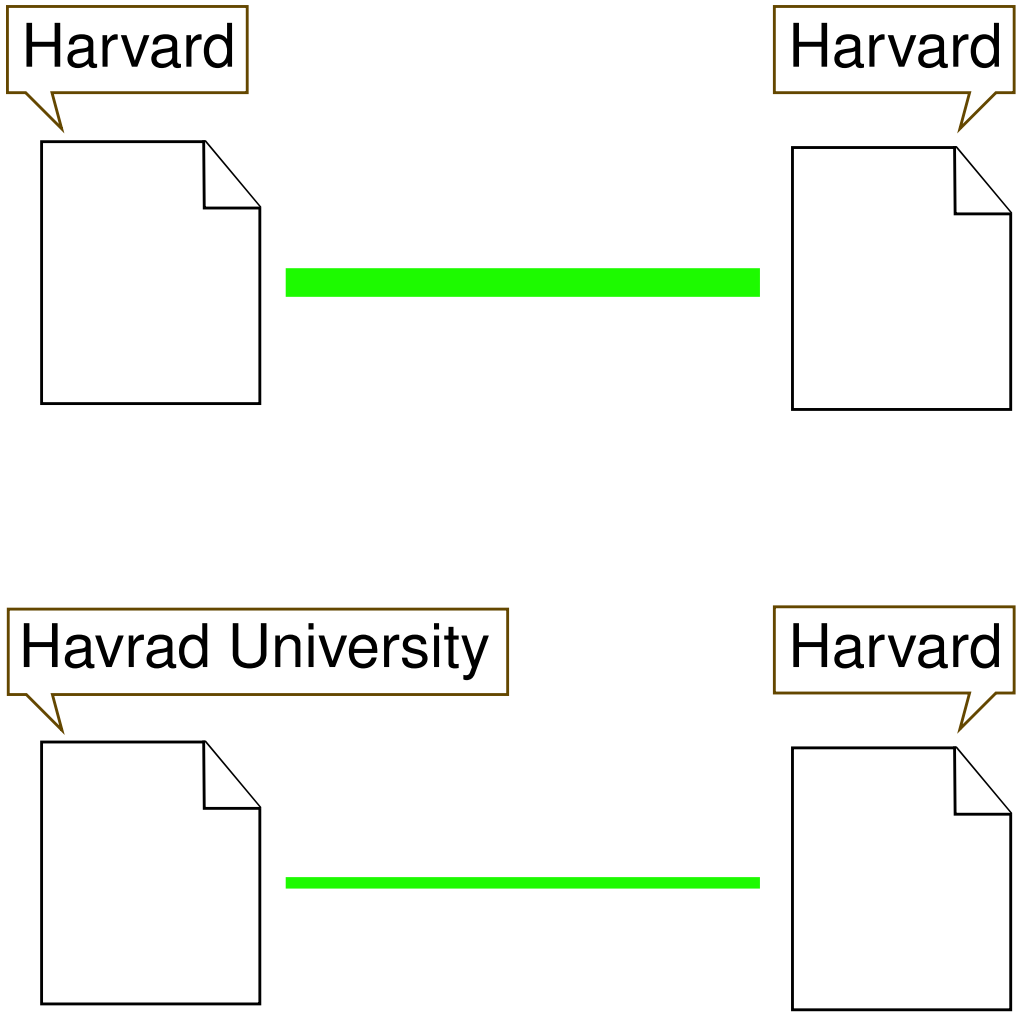

[K. Wagstaff, C. Cardie. Clustering with Instance-Level Constraints. 2000]

# Generierung von Nebenbedingungen




[K. Wagstaff, C. Cardie. Clustering with Instance-Level Constraints. 2000]

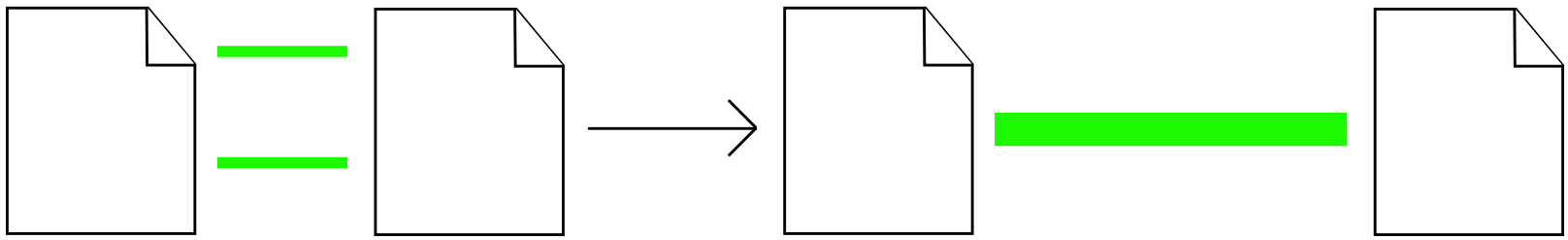
# Abgleich von Attributwerten



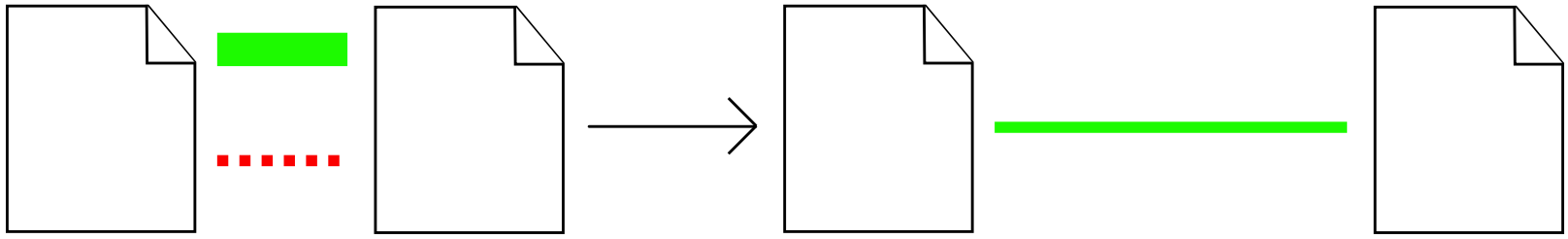

[K. Wagstaff. Intelligent Clustering with Instance-Level Constraints. 2002]

# Addition der Nebenbedingungen

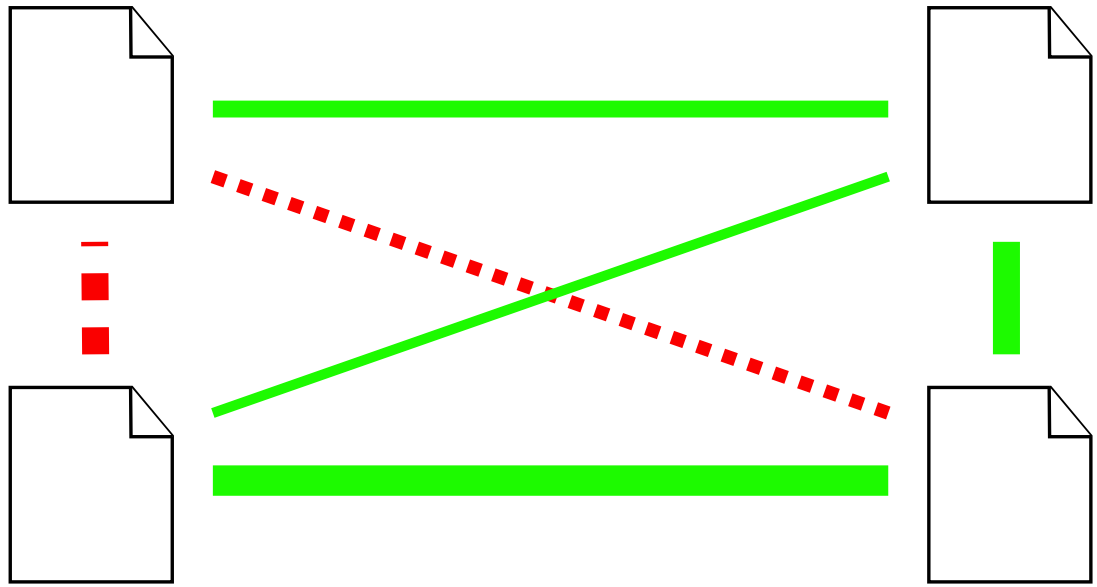
- Verstärkung



- Abschwächung

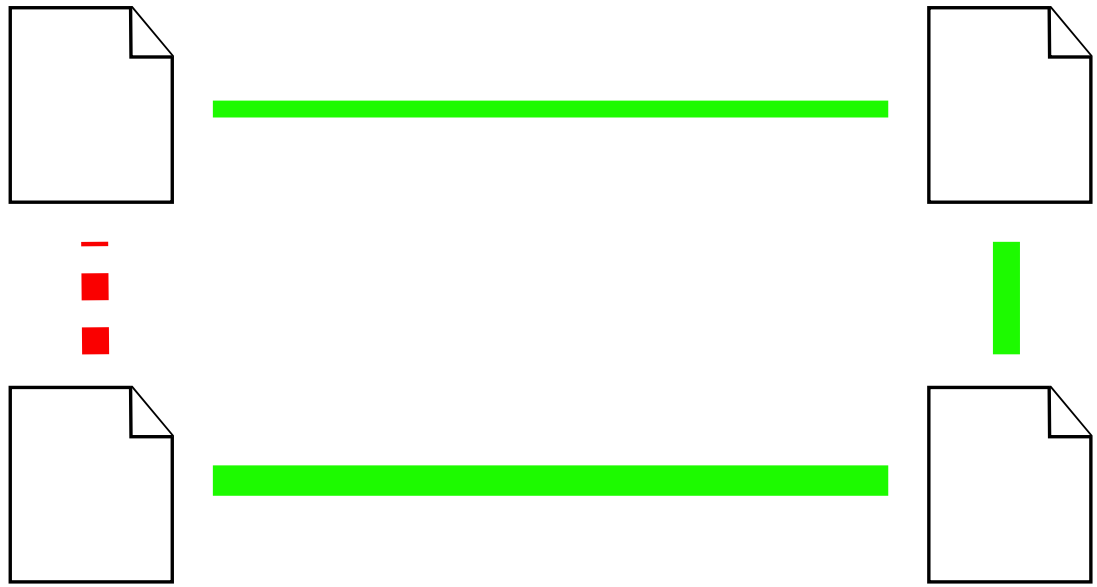


# Transitivität und Konflikte



1. Anwendung eines Schwellwertes
2. Konfliktbehebung

# Transitivität und Konflikte



1. Anwendung eines Schwellwertes
2. Konfliktbehebung

# Transitivität und Konflikte

1. Anwendung eines Schwellwertes
2. Konfliktbehebung

# Zusammenfassung: Generierung und Anwendung von Nebenbedingungen

- Abgleich von Attributwerten
- Addition von Nebenbedingungen
- Konfliktbehebung
- Gruppierung (Constrained Clustering)

# Verwendeter Korpus: WePS-2

- 30 Anfragen („*Vorname Nachname*“) an *Yahoo!*
- Je 10 Namen von 3 verschiedenen Quellen
- Manuelle Annotation (Referent, Attributwerte) und Überprüfung der erhaltenen Dokumente

| Teil des Korpus | \|Dokumente\| | \|Referenten\| |
|---|---|---|
| Englische Wikipedia | **94,0** | 10,7 |
| ACL'08 | 81,6 | 14,2 |
| 1990 US Zensus | 80,2 | **30,3** |
| Gesamter Korpus | 85,3 | 18,4 |

Durchschnittswerte für eine Anfrage

[J. Artiles et al.. WePS-2 Evaluation Campaign: Overview of the Web People Search Clustering Task. 2009]

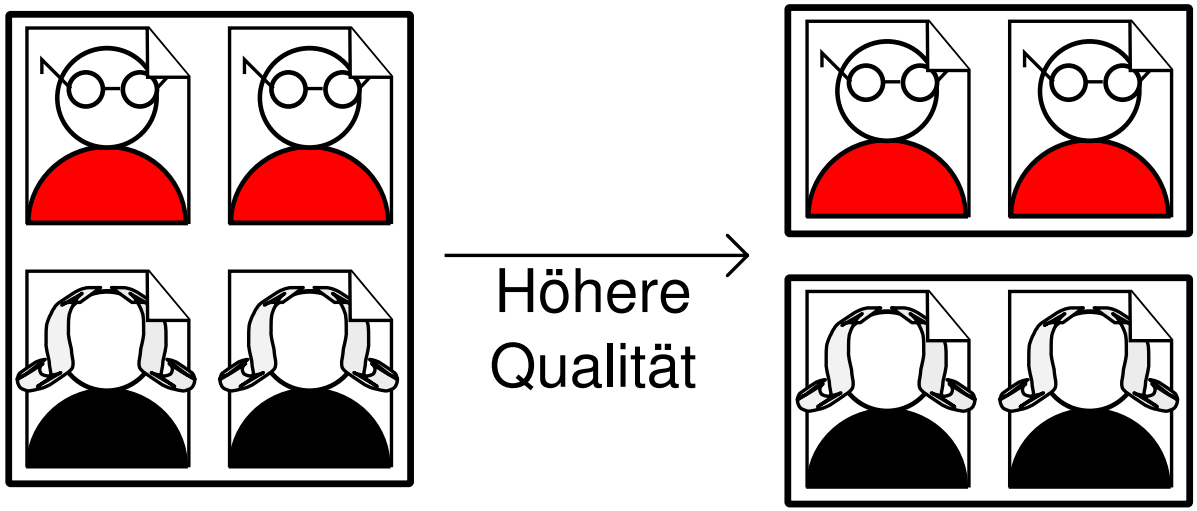
# Verwendeter Korpus: WePS-2 (Personenattribute)

| Attribut | Werte per Dokument | Werte per Referent | Referenten per Wert |
|---|---|---|---|
| Beruf | **1,07** | **9,60** | 1,04 |
| Zugehörigkeit | 1,04 | 8,36 | 1,00 |
| Schule | 0,17 | 2,81 | 1,01 |
| Geburtsort | 0,10 | 2,44 | 1,00 |
| Geburtstag | 0,10 | 1,11 | 1,05 |
| 2. Vorname | 0,09 | 1,02 | **1,07** |
| Nationalität | 0,08 | 1,18 | 1,00 |
| E-Mail | 0,07 | 1,03 | 1,01 |

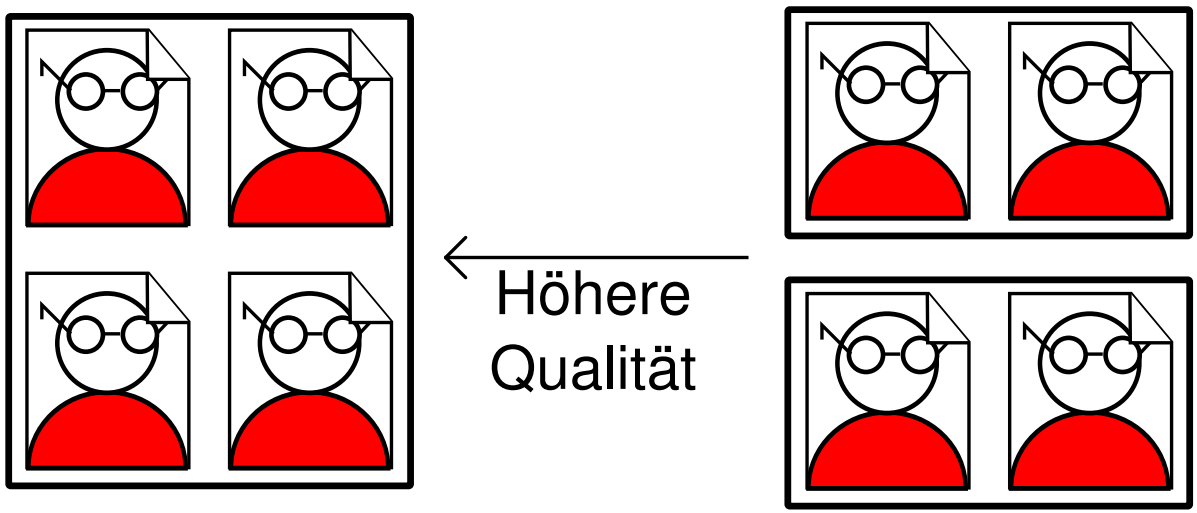
Durchschnittswerte für eine Anfrage

# Evaluation des Clusterings: BCubed $F_{\alpha=0,5}$-Measure

## Homogenität

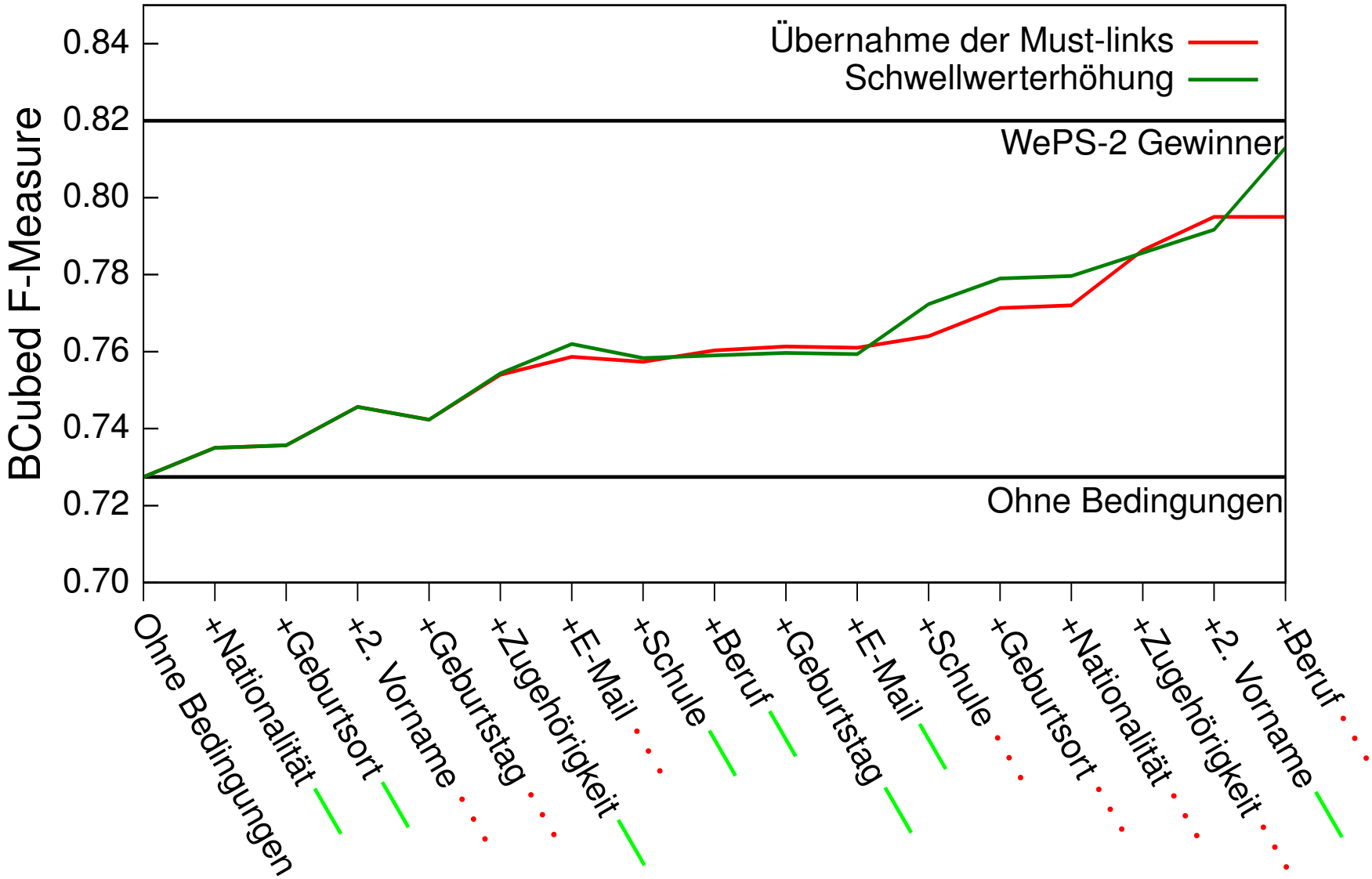



## Vollständigkeit




[E. Amigó et al.. A Comparison of Extrensic Clustering Evaluation Metrics based on Formal Constraints. 2009]

# Auswirkungen der Nebenbedingungen



# Zusammenfassung

- Anwendung von *Constrained Clustering* in der Personensuche im Internet
- Verwendung von unsicheren Informationen von verschiedenen Personenattributen
- Vorschlag eines generischen Systems zur Generierung von Nebenbedingungen

| Verfahren | $F_{\alpha=0,5}$ |
|---|---|
| PolyUHK | 0,82 |
| **BUW** | 0,81 |
| UVA | 0,81 |
| ITC-UT | 0,81 |
| XMEDIA | 0,72 |
| UCI | 0,71 |

# Ausblick

- Anwendung der gewichteten Nebenbedingungen zur Veränderung der Dokumentähnlichkeit
- Testen verschiedener Methoden zur automatischen Bestimmung von Schwellwerten
- Genauere Analyse der verschiedenen Kombinationen von Personenattributen

# Ausblick

- Anwendung der gewichteten Nebenbedingungen zur Veränderung der Dokumentähnlichkeit
- Testen verschiedener Methoden zur automatischen Bestimmung von Schwellwerten
- Genauere Analyse der verschiedenen Kombinationen von Personenattributen

Vielen Dank für Ihre Aufmerksamkeit