Exploratory Search Pipes with Scoped Facets

Tim Gollub   Leon Hutans   Tanveer Al Jami   Benno Stein

Bauhaus-Universität Weimar
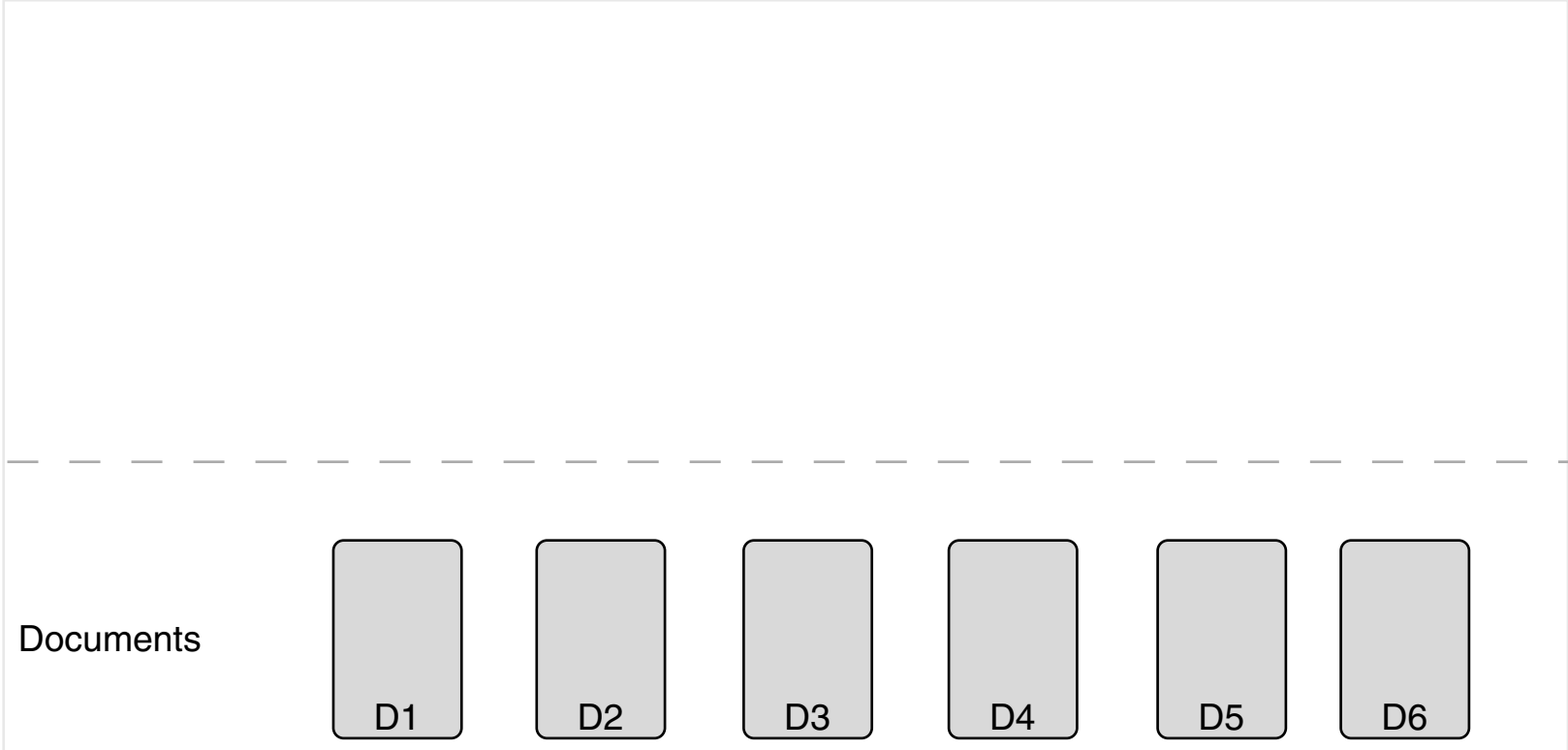
webis.de

Research Context
DFG-Project Process-Oriented Discourse Analysis

❑ Digital Humanities Project.

❑ We develop *distant reading* devices.

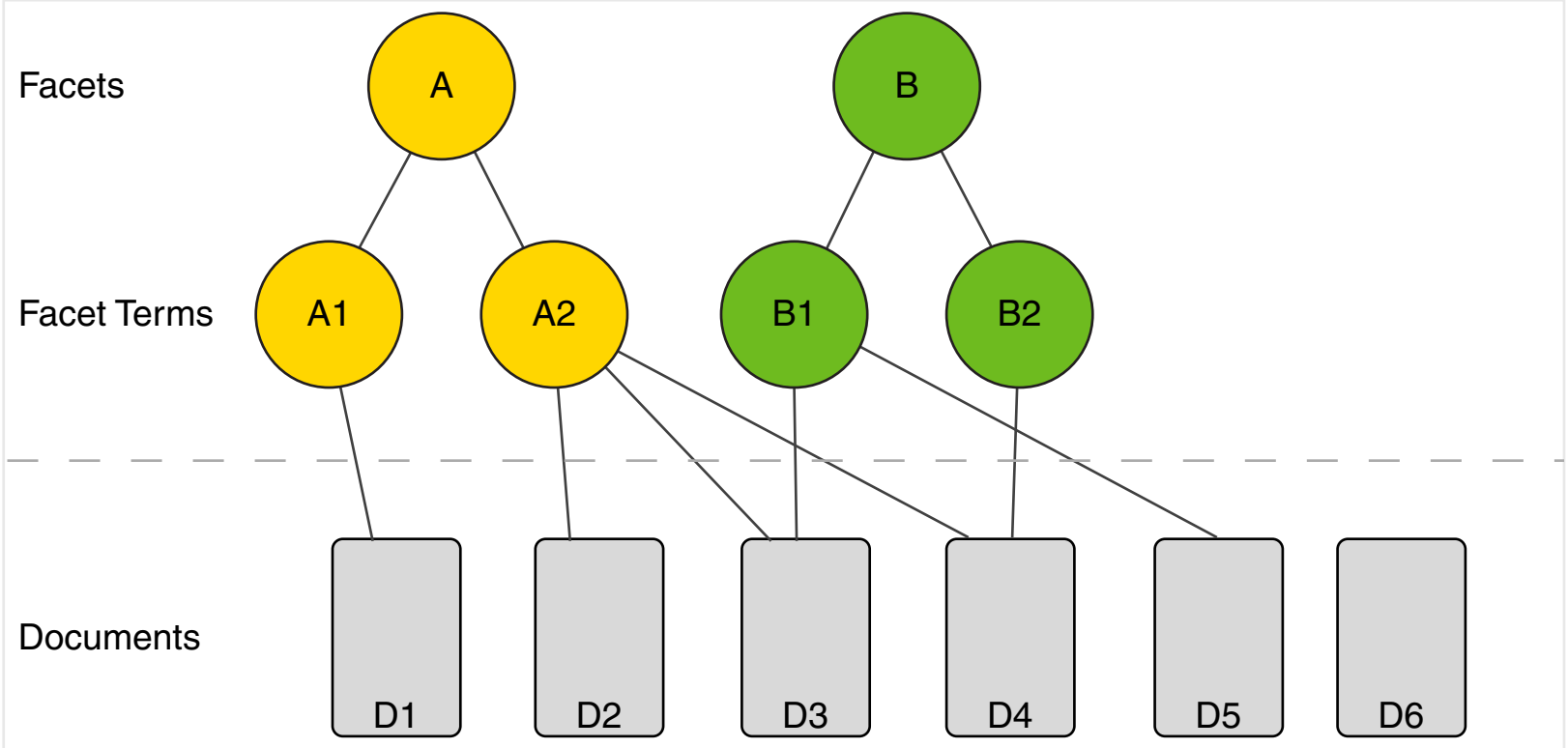❑ First device is a *faceted search system* for *content-level* explorations of document collections.

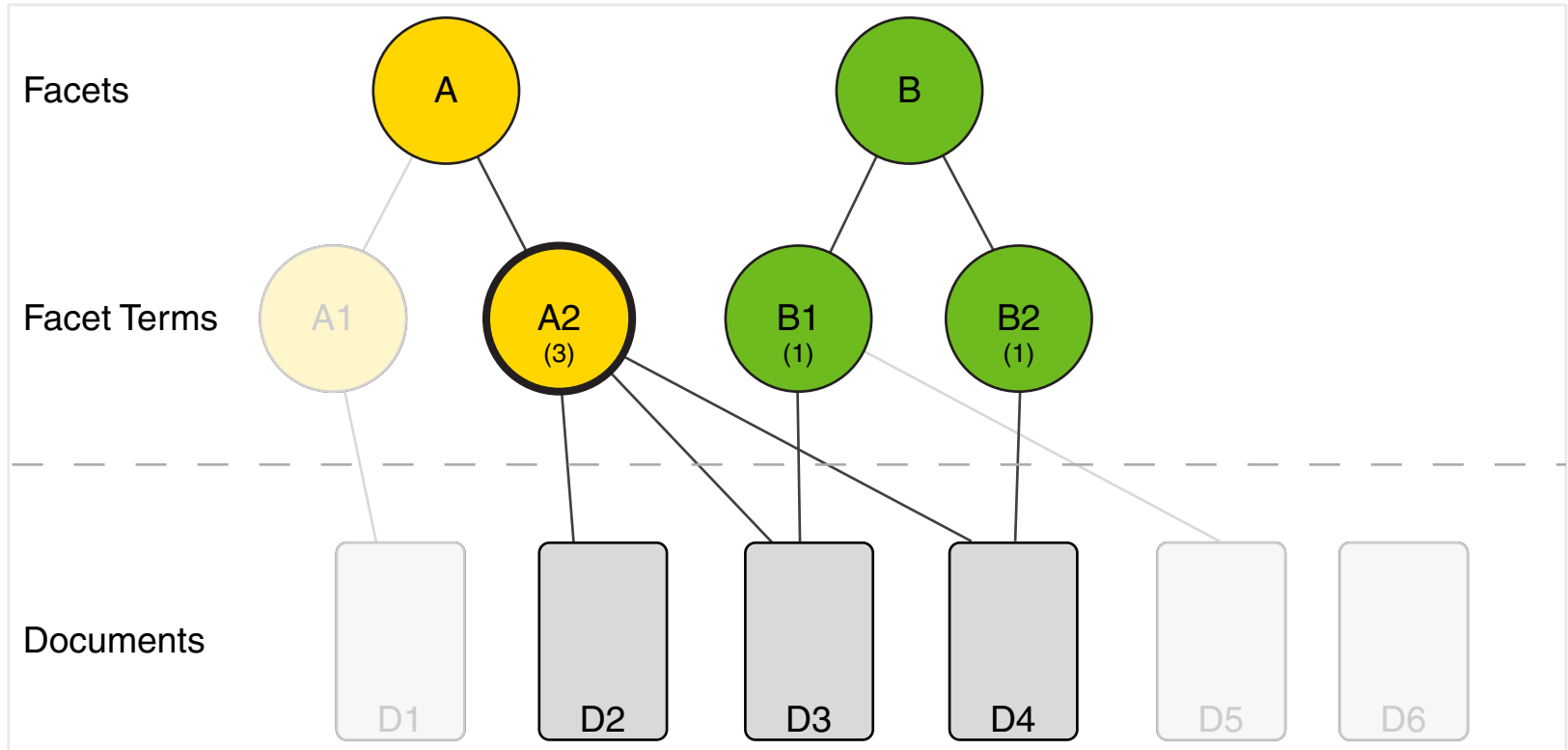# Classical Facet Search Systems

Documents

Documents

D1	D2	D3	D4	D5	D6

Classical Facet Search Systems

Facets

Facets: A (yellow), B (green)

Facet Terms: A1, A2 (yellow), B1, B2 (green)

Documents: D1, D2, D3, D4, D5, D6
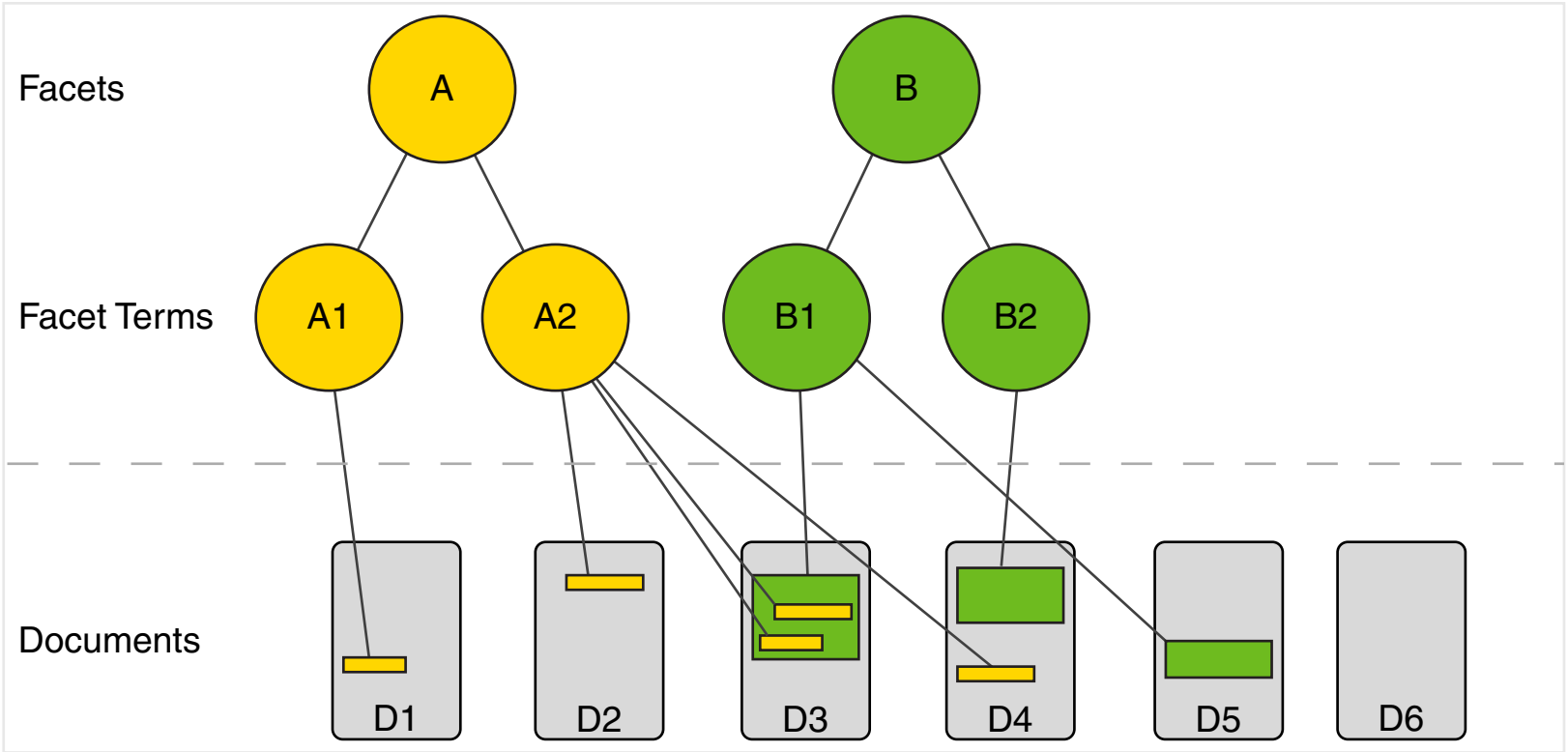
Classical Facet Search Systems
Filtering on User Selection

❑ Issue: Documents as atomic context units are not flexible enough to accommodate for content-based explorations.
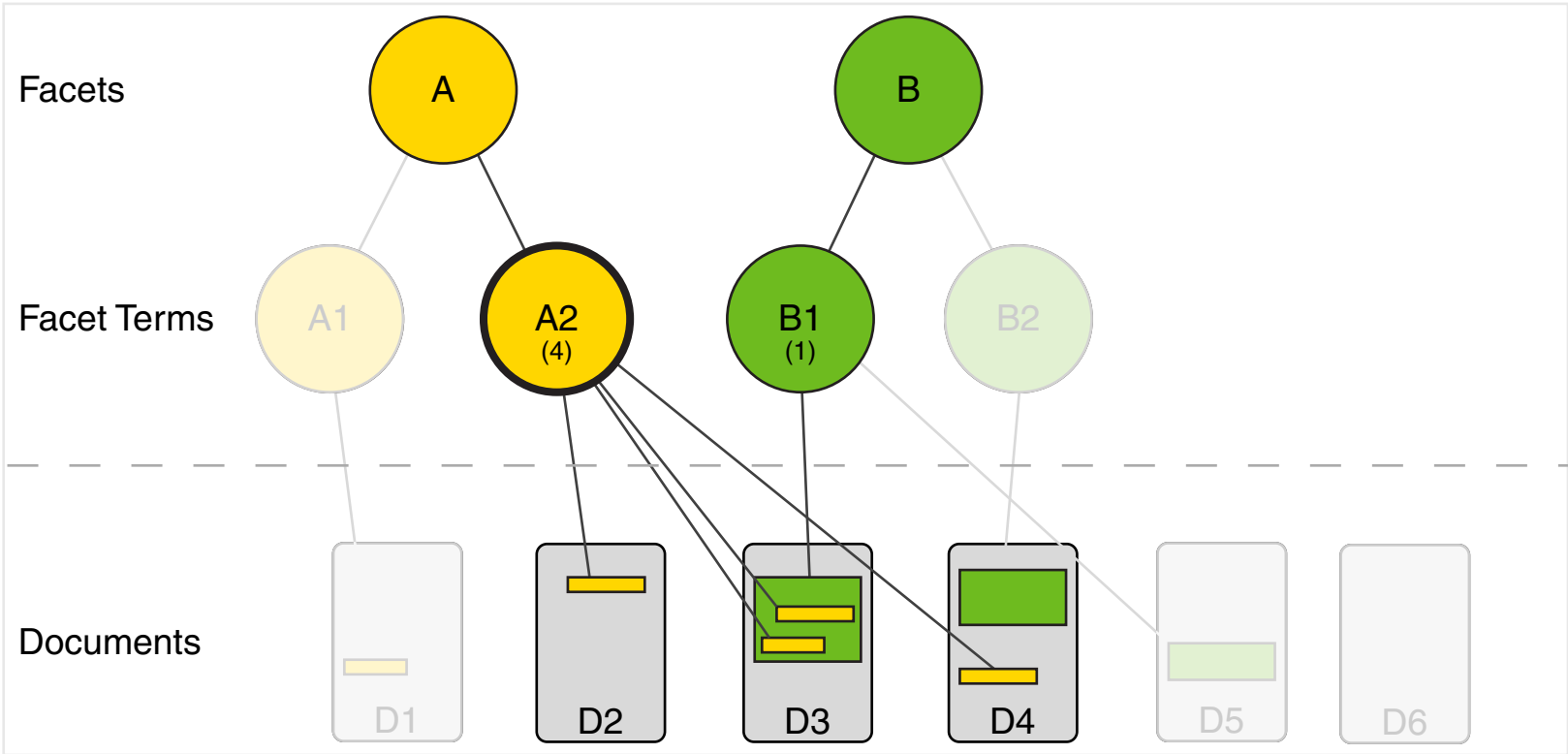
Span-Based Faceted Search
Annotation of Character Spans

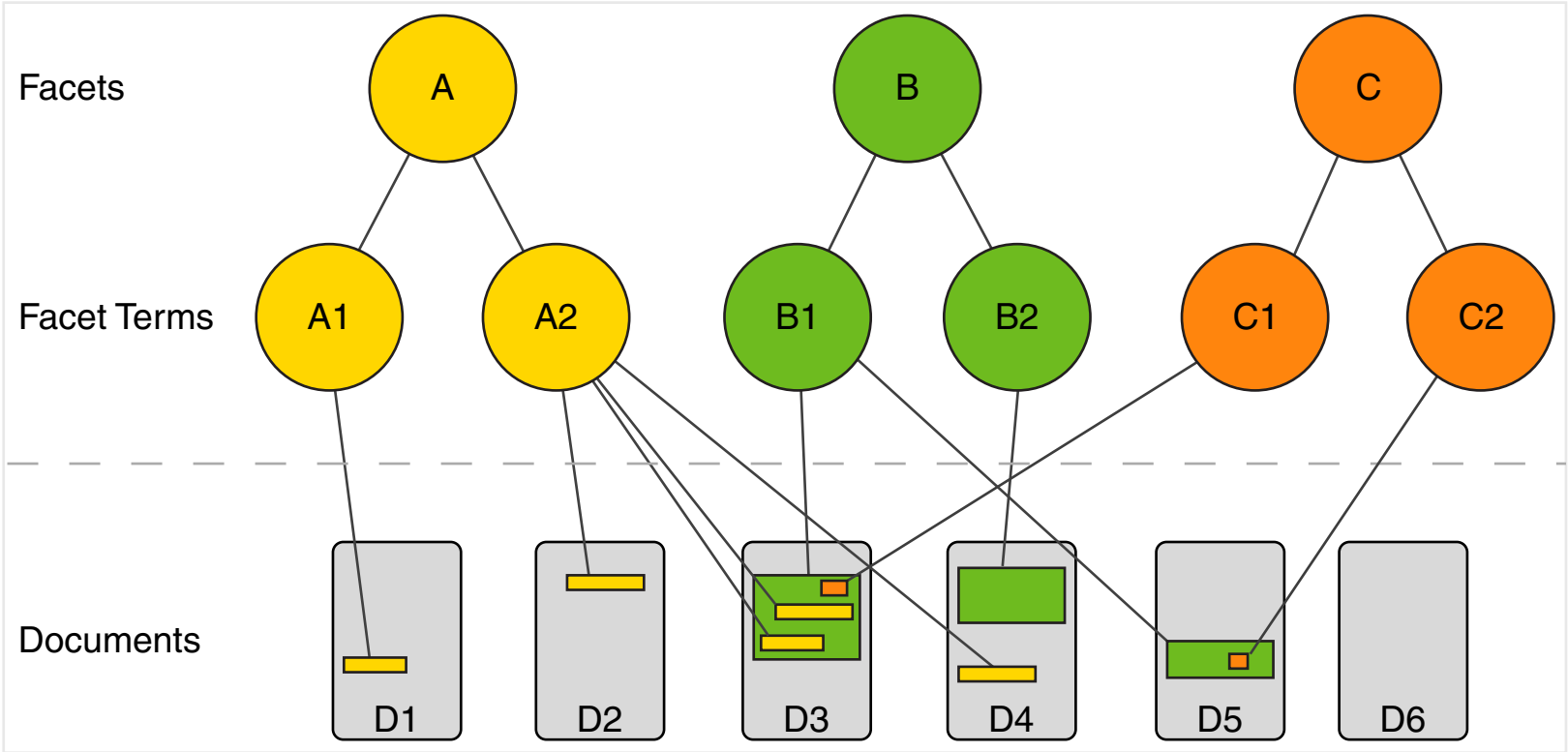# Span-Based Faceted Search
Filtering on User Selection

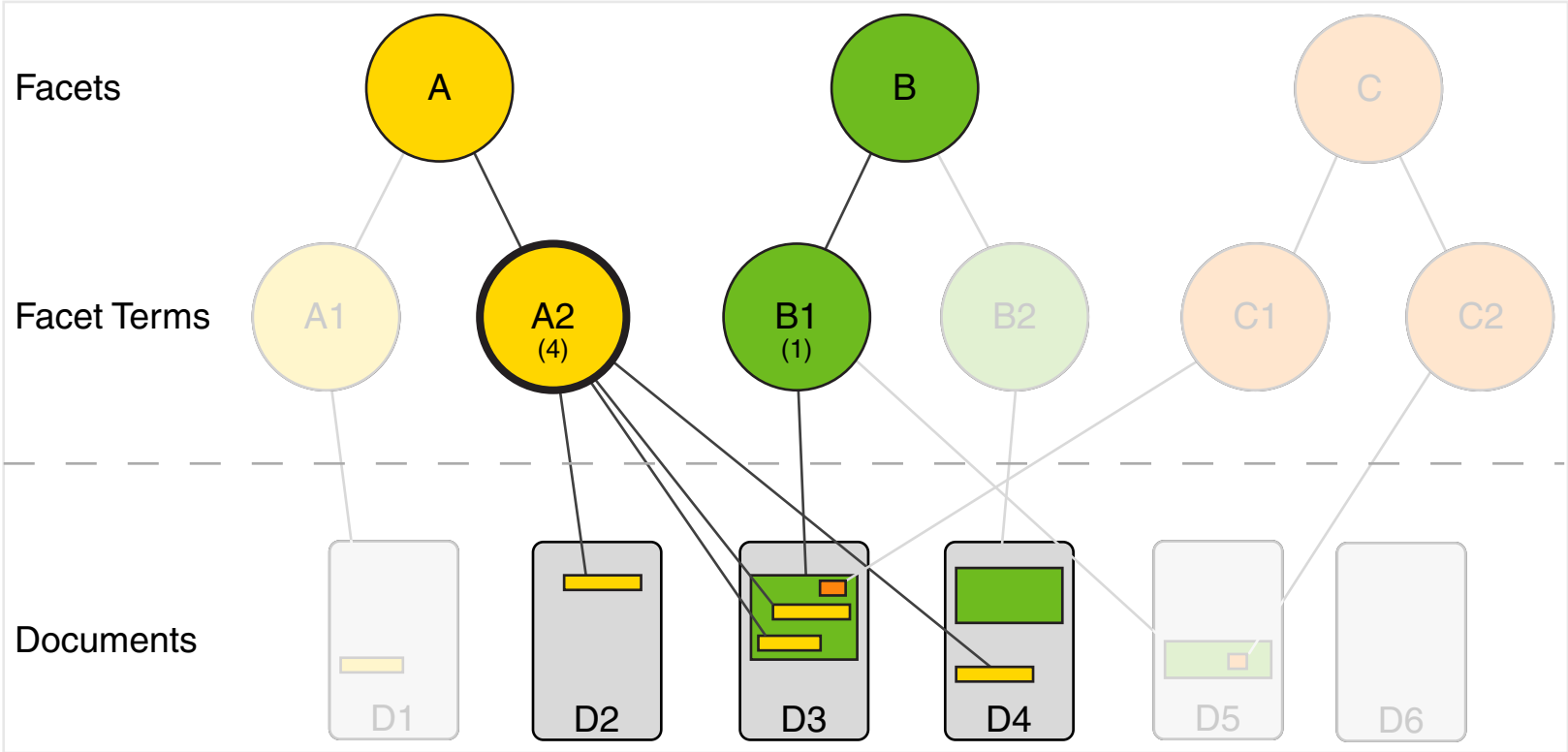❑ Benefit: Evaluation of facets within different contexts is now possible.

Span-Based Faceted Search

Three Facets Case

Span-Based Faceted Search
Filtering on User Selection

- Issue: A common use case is to search for facet terms (e.g. entities) that occur in a common context. This cannot be done with global intersections.

Exploratory Search Pipes
Sequential Evaluation of Facets
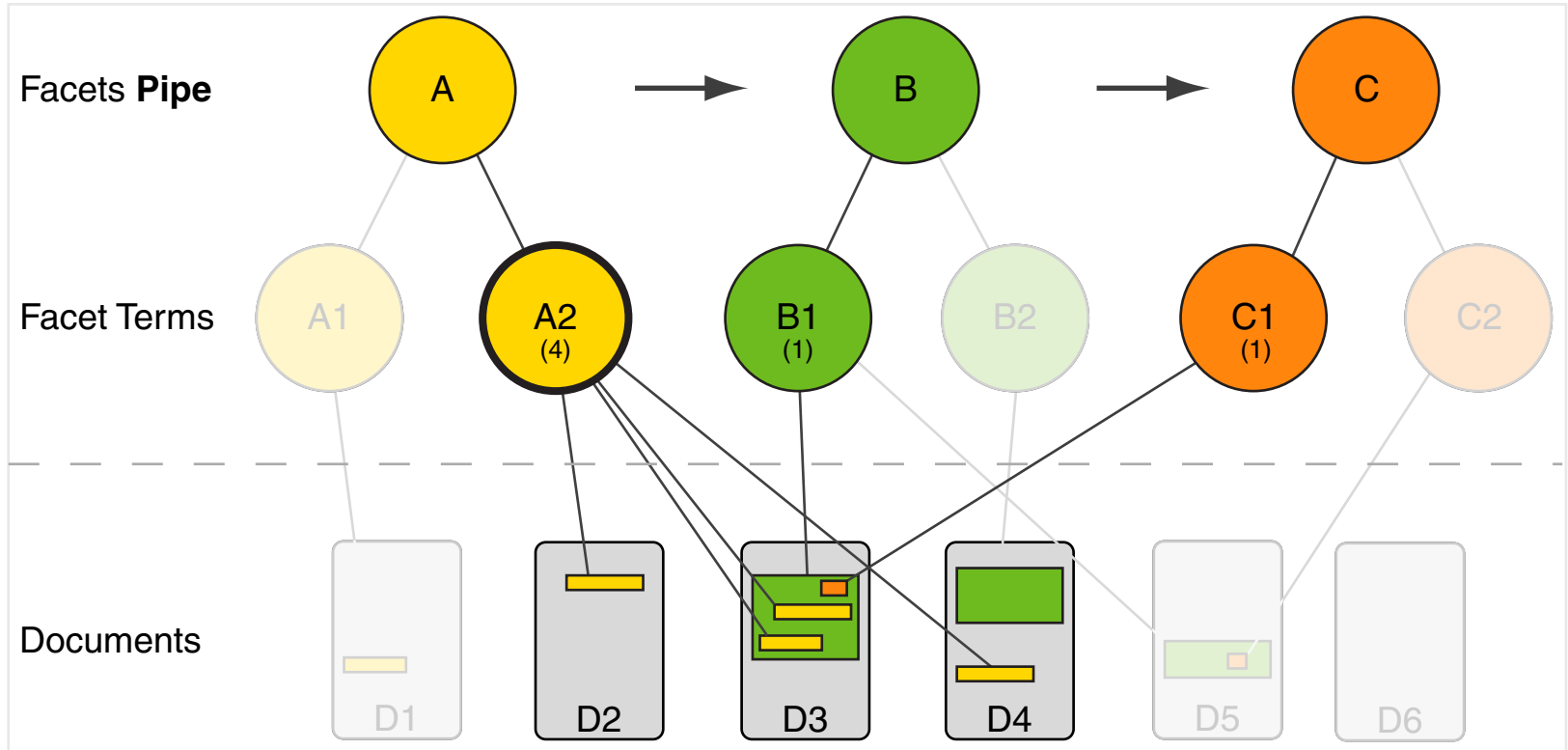
❑ Benefit: Facets are evaluated sequentially one after the other. The result set for a facet is determined from intersections with the character spans of the facets terms from the facet to the left (if any).

Scoped Facets
Impact of Facet Scope on Result
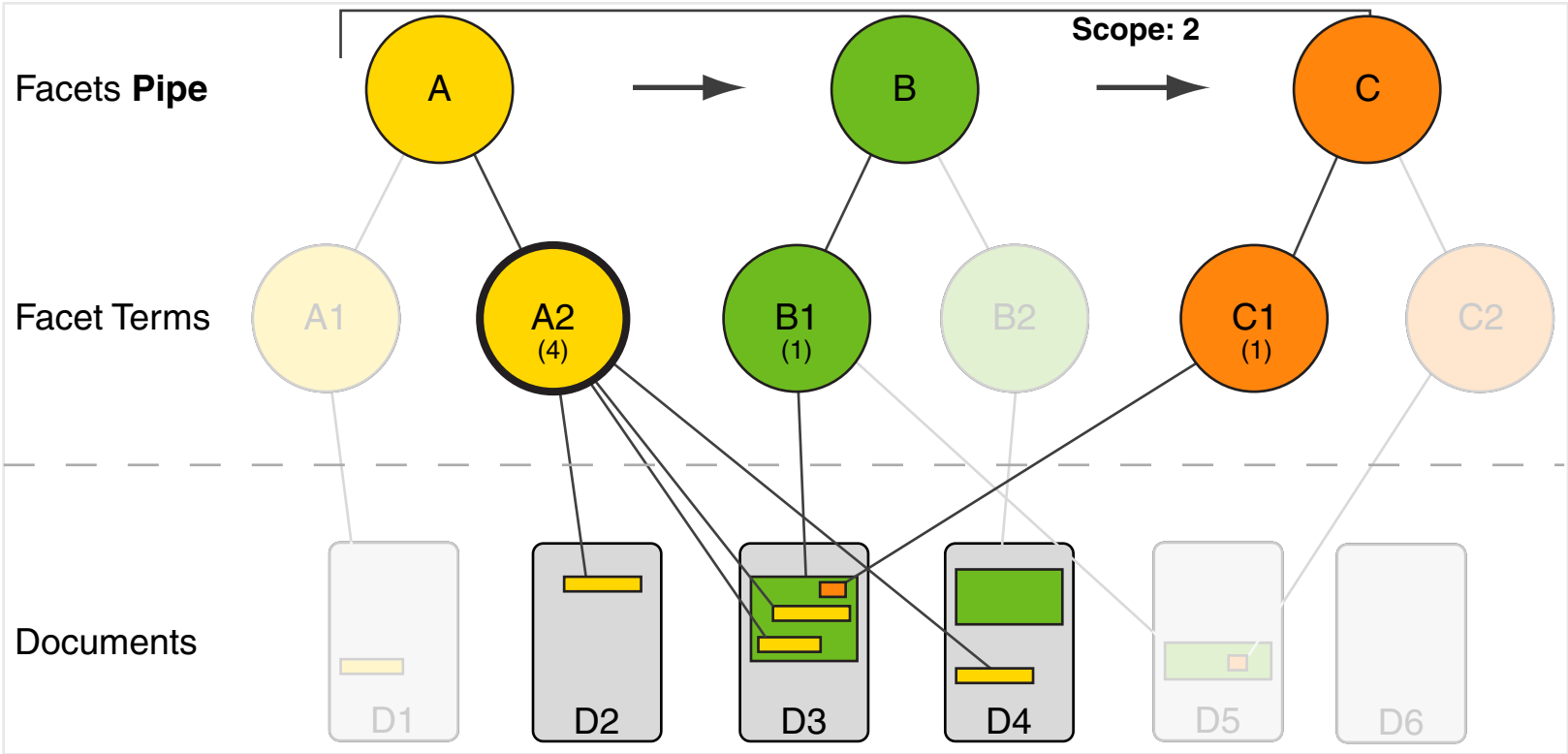
- Issue: A common use case is to filter the facet terms of facet B with respect to A2, but then to use all spans of the remaining facet terms in Facet B (=B1) for the intersections with Facet C.

Scoped Facets

Impact of Facet Scope on Result
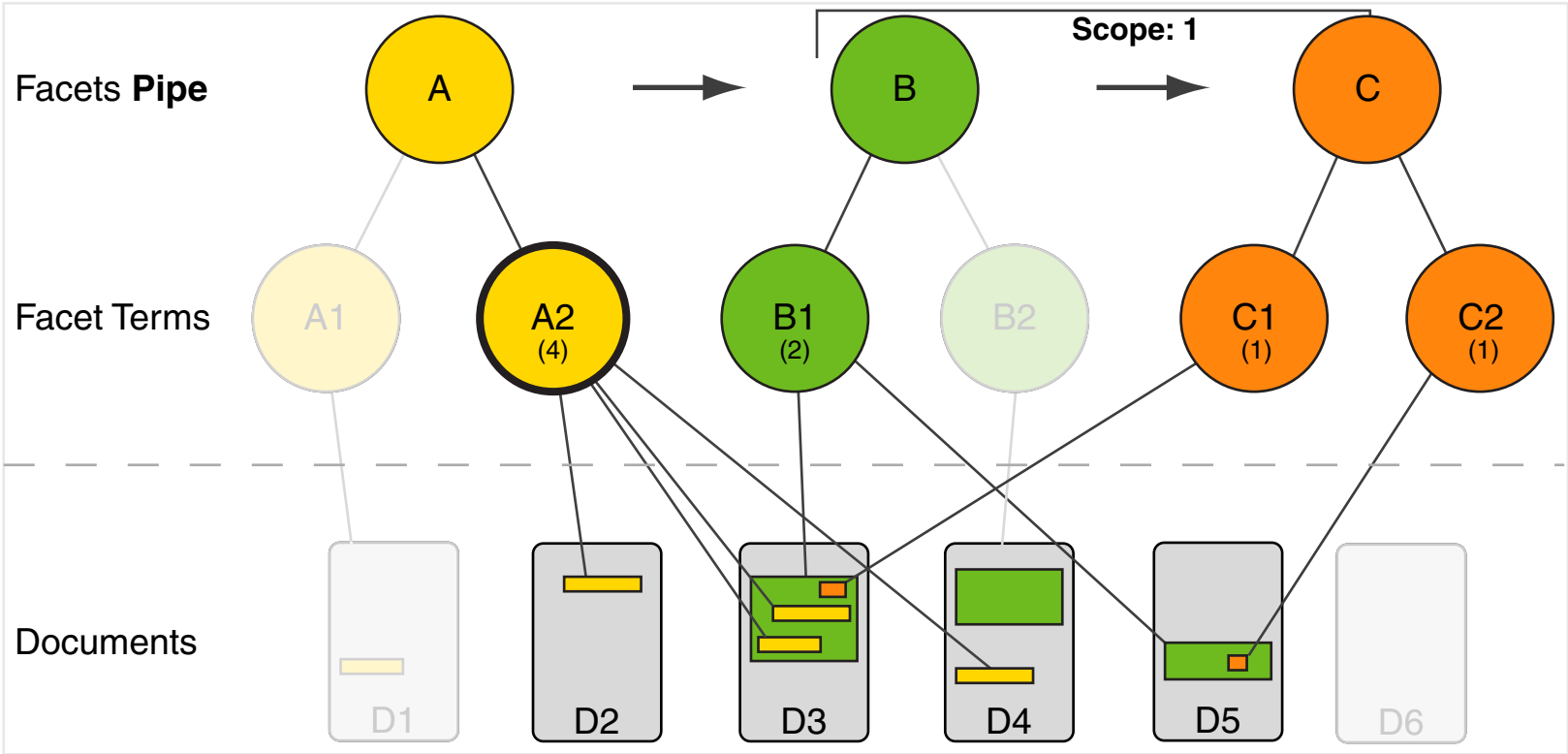
- Benefit: With the introduction of facet scopes, the user can decide which character spans of a facet term to take into account.

Prototype Podascope

Facet Navigation and Search

Facets (15)

Scope: 0

| / | Search | + |

Total 15 results

Authors (70601)

Company (282)

Conferences (47)

Country (65)

Database (5)

Disease (56)

Language (96)

Location (325)

Organisation (708)

Person (209)

Programming Language (39)

Prototype Podascope

Facet Navigation and Search

Conferences (47)

Scope: 0

| / | Conferences | Search | ➕ |

Total 47 results

AAAI

AAMAS

ACL

ACL (1)

ACL (2)

ACM DL

ADCS

AIRS

BIR@ECIR

Challenged Networks

CHIIR

Prototype Podascope
Facet Navigation and Search

Conferences (1/47)

Scope: 0

/	Conferences	ictir

Total 1/47 results

ICTIR

Prototype Podascope
Pipe Construction

Conferences (1/47) **Facets (10)**

Scope: -1 ○────────────────────

← | / | Search | ⊕

Total 10 results

1	Authors (528/70601)
1	Company (4/282)
1	Conferences (1/47)
1	Country (1/65)
1	Location (2/325)
1	Organisation (7/708)
1	Person (3/209)
1	Publications (314/64287)
1	Titles (312/64079)
1	Years (7/47)

Prototype Podascope

Pipe Construction

Conferences (1/47)   **Publications** (314)

Scope: -1 ○━━━━━━━━━━━━

← / Publications | Search | ＋

Total 314 results

1	"A term is known by the company it keeps": On Selecting a Good Expansion Set in Pseudo-Relevance Feedback.
1	<i>Merge-Tie-Judge</i>: Low-Cost Preference Judgments with Ties.
1	A Belief Model of Query Difficulty That Uses Subjective Logic.
1	A Collaborative Ranking Model with Multiple Location-based Similarities for Venue Suggestion.
1	A Comparative Study of Pseudo Relevance Feedback for Ad-hoc Retrieval.
1	A Contextual Bandit Approach to Dynamic Search.
1	A Descriptive Approach to Classification.
1	A Diagnostic Study of Search Result Diversification Methods.
1	A Formal Account of Effectiveness Evaluation and Ranking Fusion.

Prototype Podascope

Pipe Construction

Conferences (1/47)   **Publications** (314)   **Authors** (528)

Scope: -2 ○

| / | Authors | Search |

Total 528 results

10	Leif Azzopardi
10	Oren Kurland
10	W. Bruce Croft
9	Guido Zuccon
8	Haggai Roitman
7	Ben Carterette
7	ChengXiang Zhai
7	Christina Lioma
7	Craig Macdonald
7	Dawei Song 0001
7	Hui Fang 0001

Prototype Podascope

Scope Reduction

Scope: -1

/ | Conferences | Search

Total 47 results

528	ICTIR
340	SIGIR
293	CIKM
260	ECIR
208	TREC
131	WSDM
98	AIRS
76	JCDL
73	AAAI
72	CHIIR
62	IIiX

User Study
Comparison with DBLP

- ❑ 14 participants with computer science background.
- ❑ Two phases: (P1) Simple search tasks. (P2) Complex search tasks.
- ❑ User Experience Questionnaire (UEQ)

UEQ Means

Legend:
- Excellent
- Good
- Above_avg
- Below_avg
- Bad
- -●- DBLP P1
- -◆- PODA P1
- ─●─ DBLP P2
- ─◆─ PODA P2
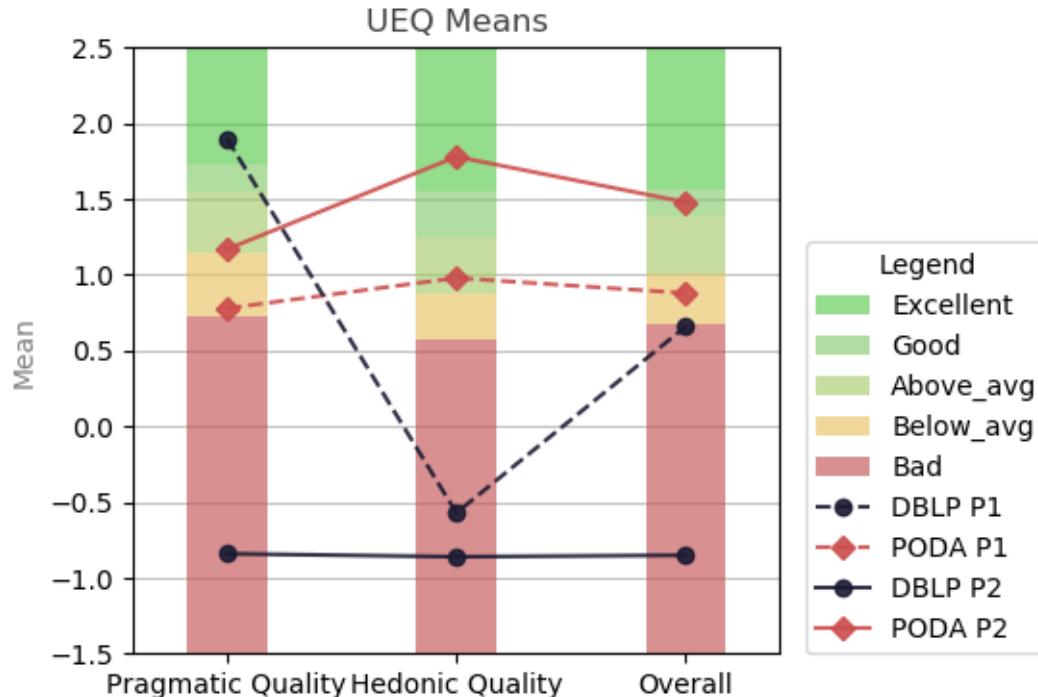
- ❑ Overall, our system PODA (red) is assessed equally to DBLP (black) after P1 (dashed), cleary superior to DBLP after P2 (solid).

Thank you!