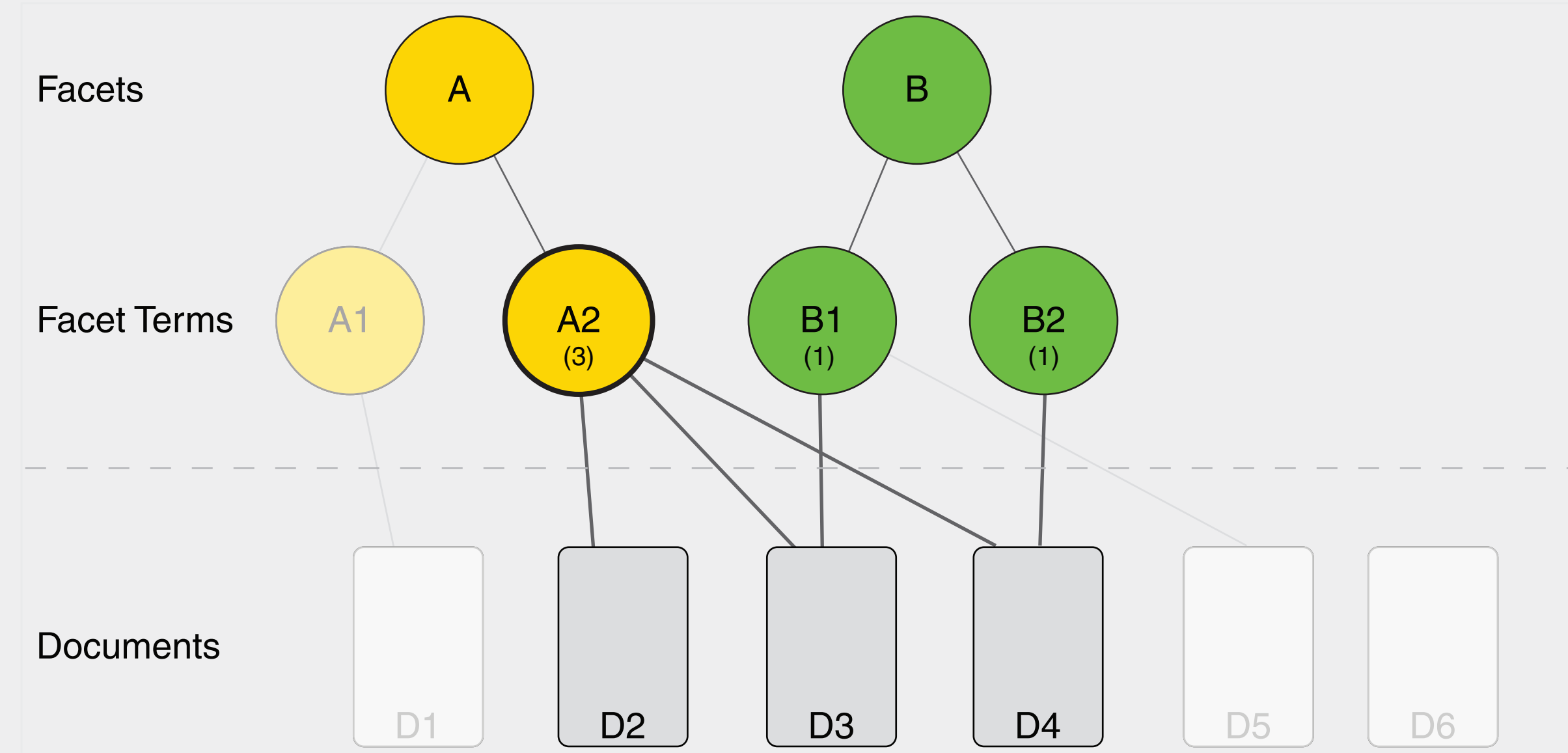


# Exploratory Search Pipes with Scoped Facets

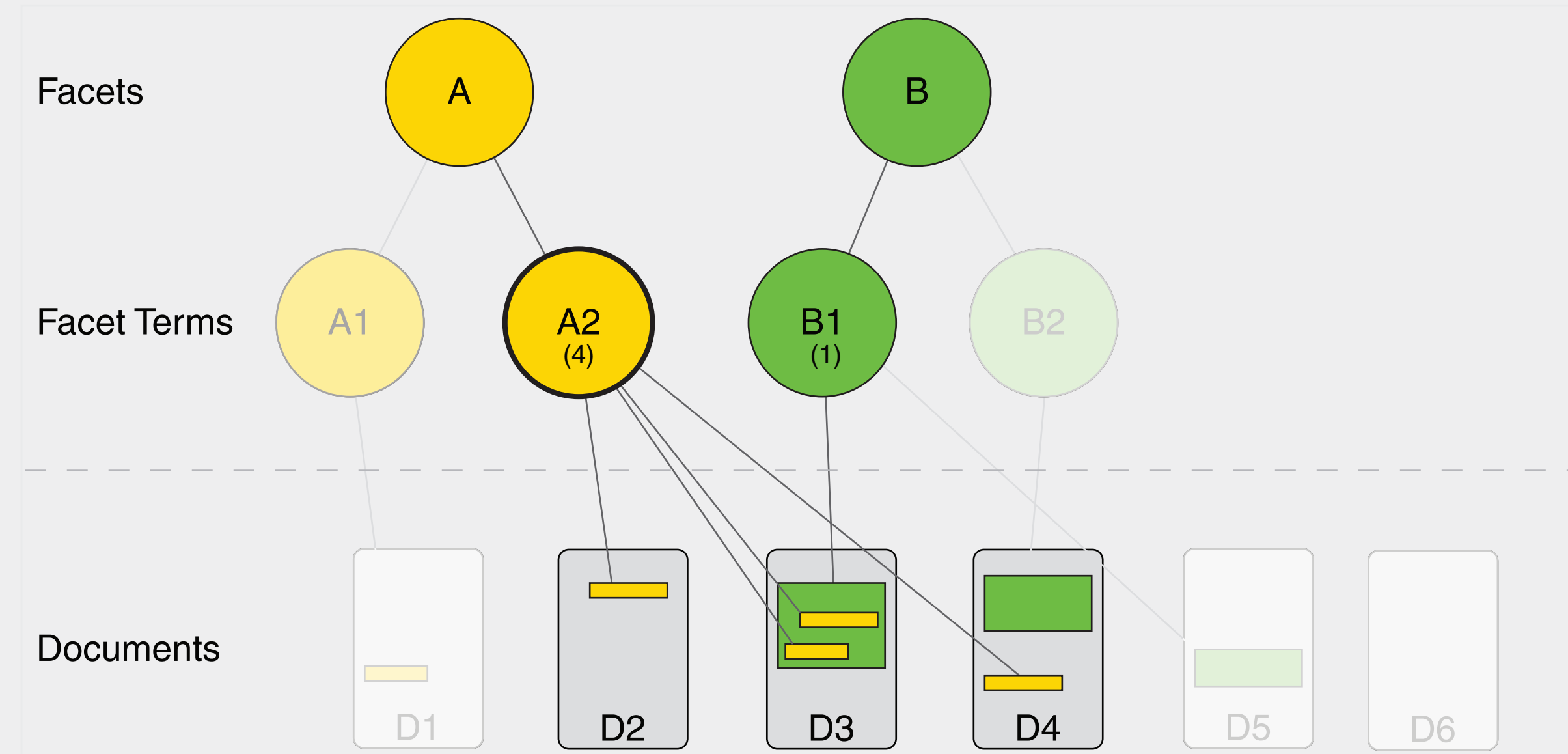
**Contribution.** We present a **faceted search system** designed for **content-level** explorations of document collections.

## Spans.

Traditional faceted search systems treat documents as atomic units.

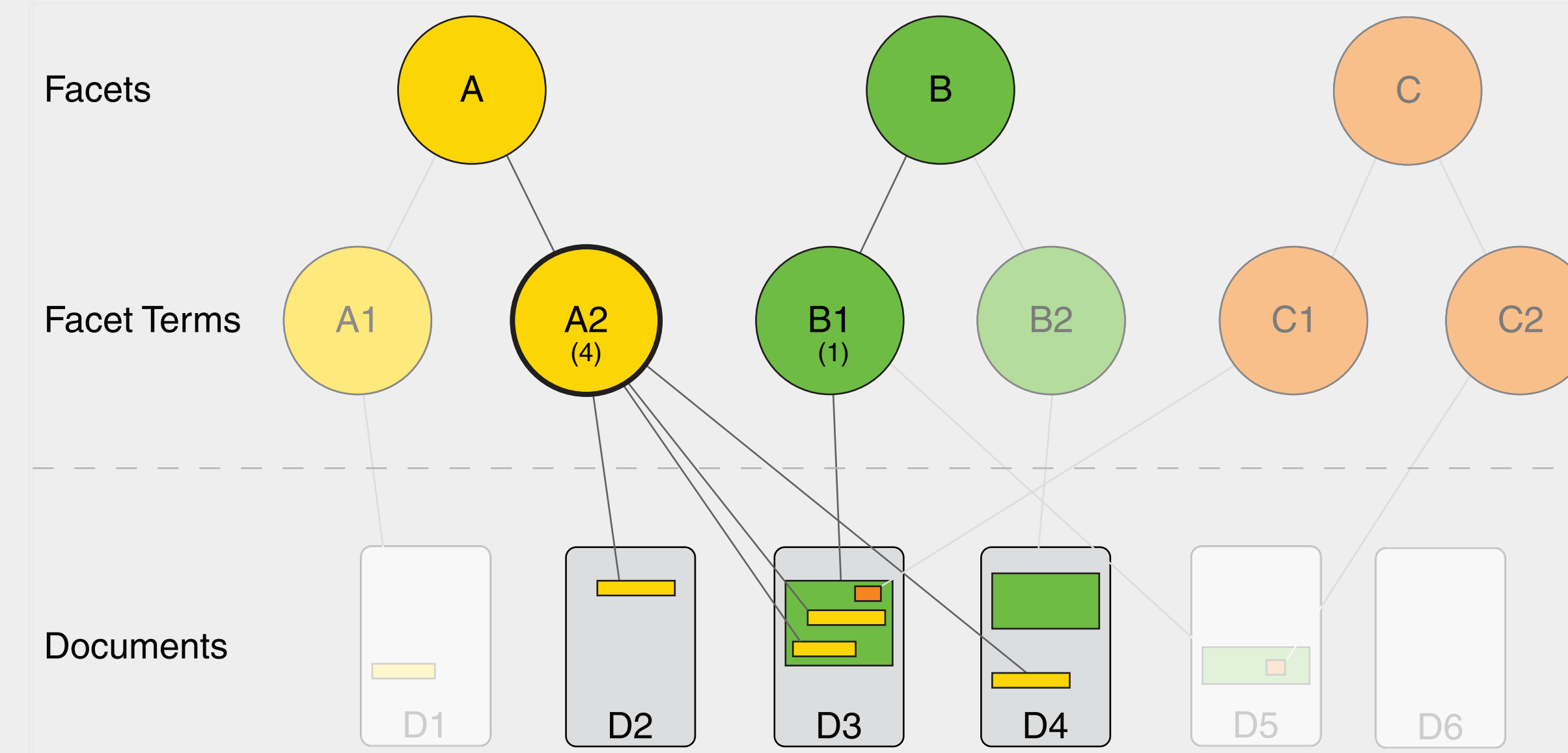


In our system, facet terms are connected to **character spans** in the documents. Two facet terms are related, only if their character spans intersect (see *Documents D3 vs. D4*).

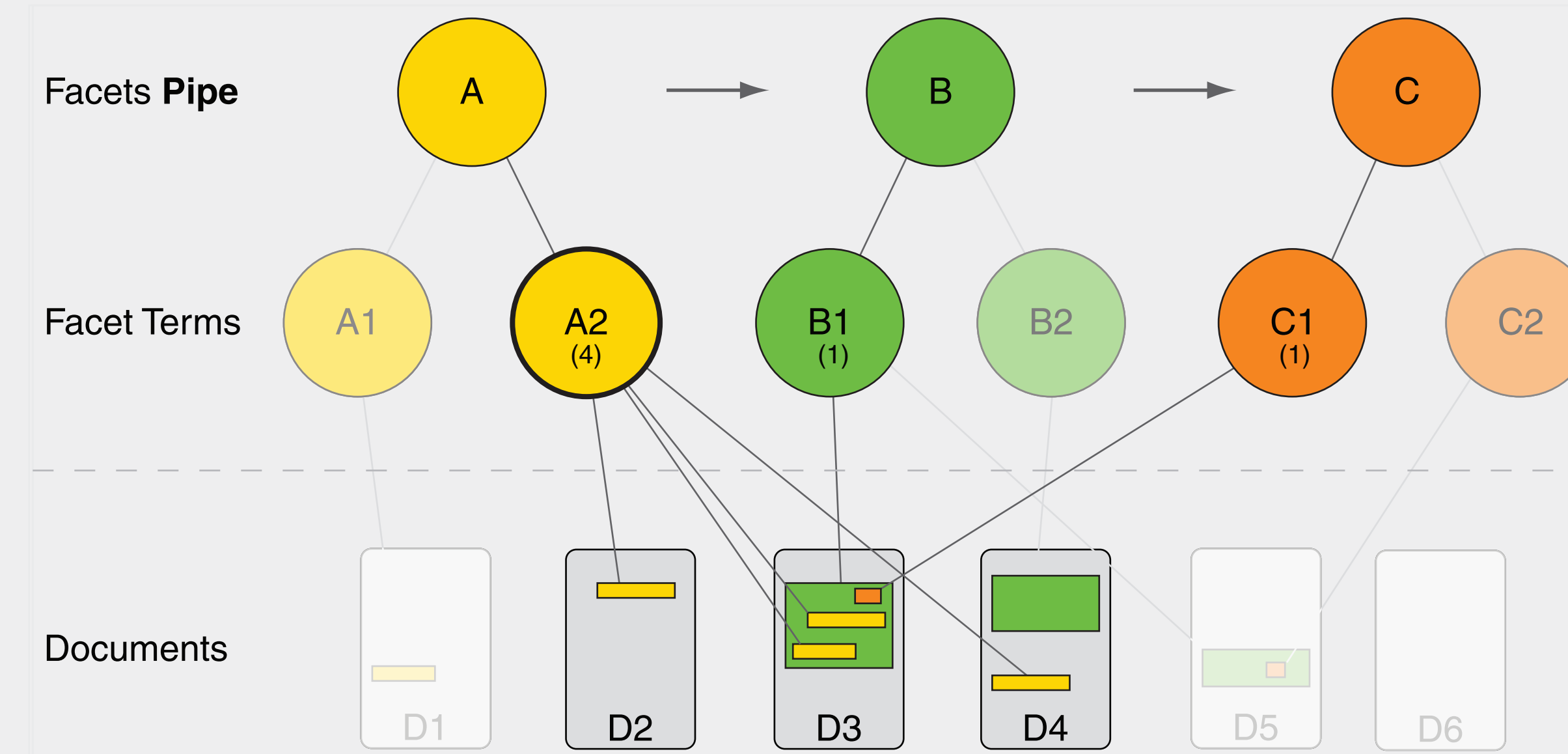


## Pipes.

With traditional facet semantics, *Facet Term C1* below is not reachable as it does not intersect with character spans of the selected *Facet Term A2*.

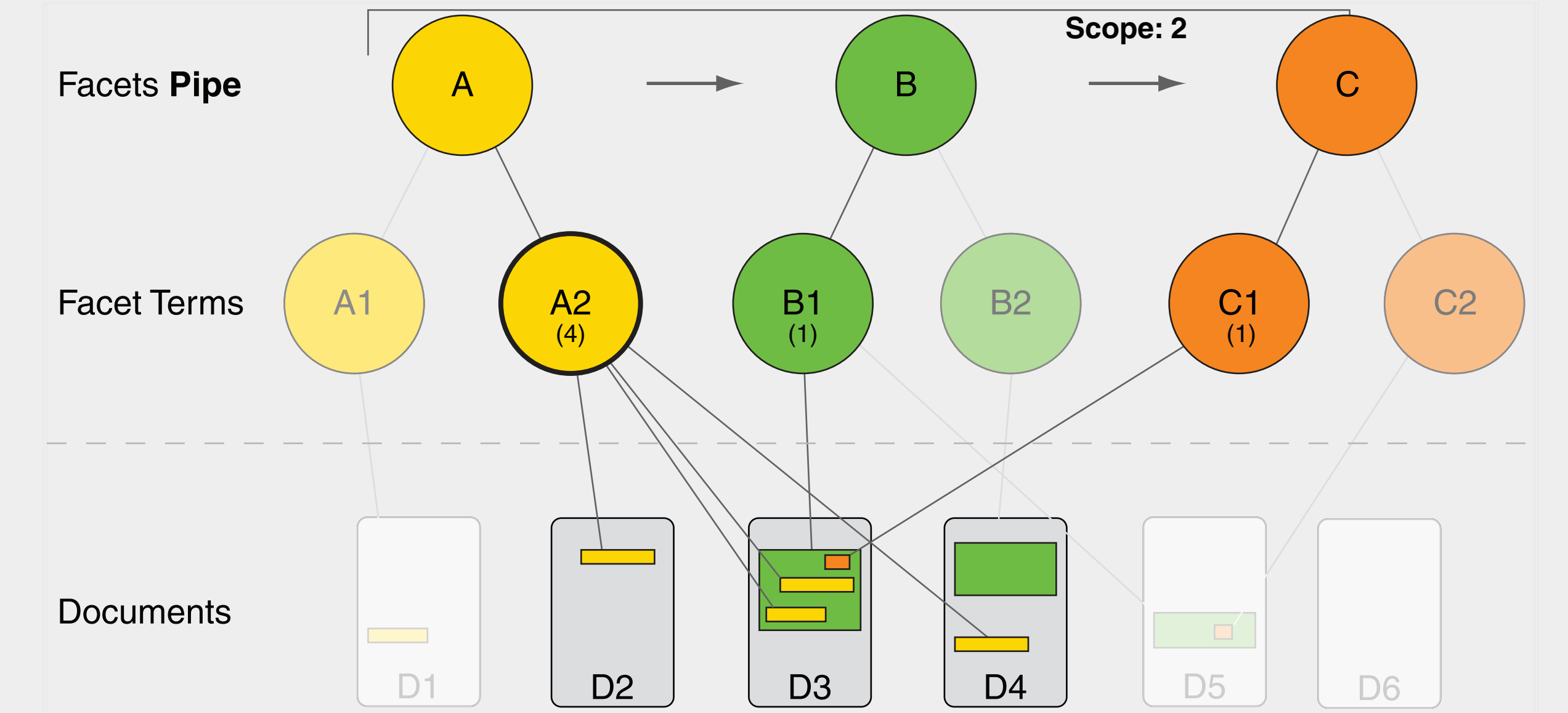


In our system, facets are **arranged as a sequence** (=pipe) by the user. The character spans of *Facet A* “activate” intersecting character spans of *Facet B*. These spans are in turn the basis to “activate” spans of *Facet C*.

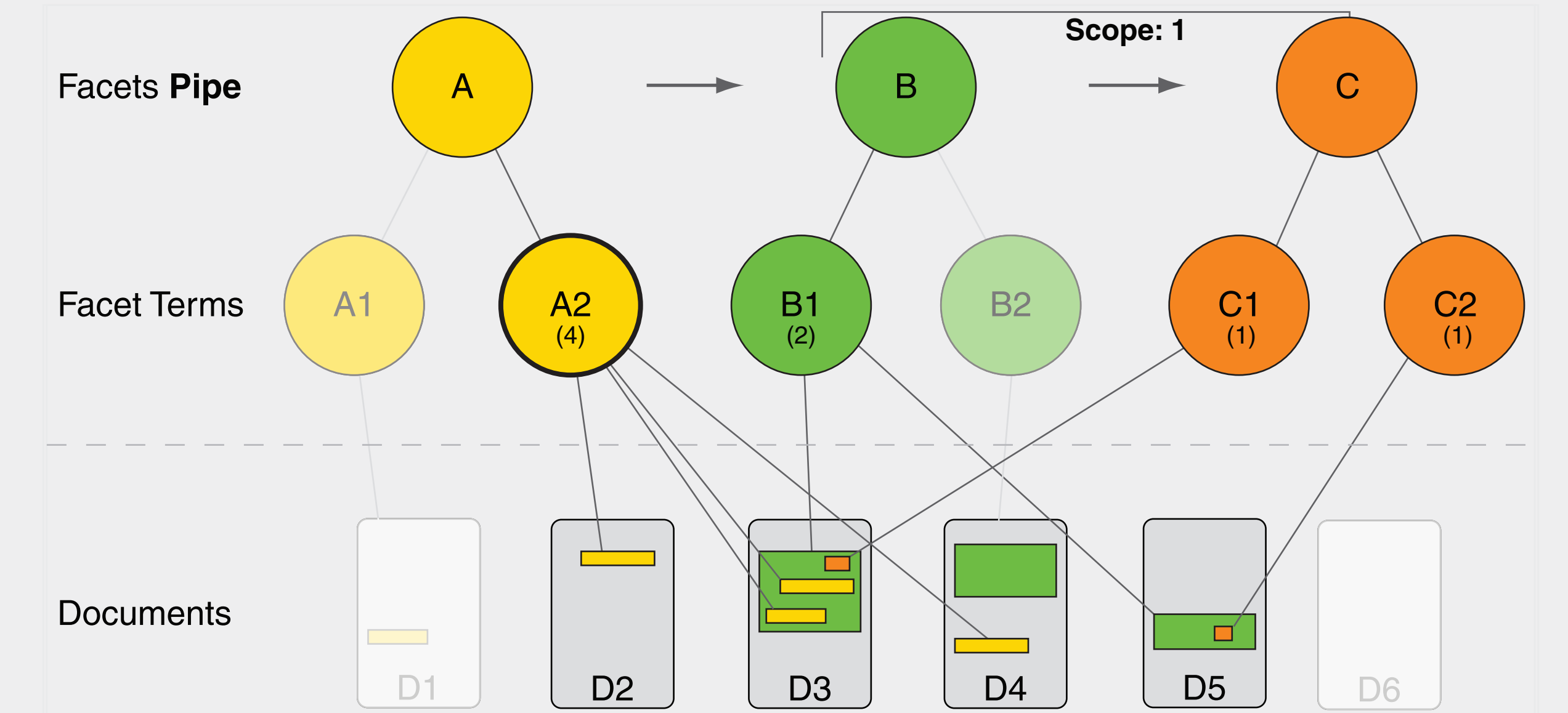


## Scopes.

With the introduction of **facet scopes**, the user can decide which prior facets should be considered to determine “activated spans”.



By **reducing the scope** of *Facet C* from two to one, also the green character span of *Facet Term B1* in *Document D5* is activated, although no character span of *Facet Term A2* intersects.



## Prototype.

Example pipe based on bibliographic records from dblp and entity linking with DBpedia Spotlight. The example e.g. reveals in Facet 3 *Conferences* similar to *ICTIR* in terms of *Author* overlap, as well as in Facet 5 *Companies* that occur in the *Publications* of *ICTIR Authors* at *IR Conferences*.

Facet	Item	Count
Authors (528)	W. Bruce Croft	47
	Arjen P. de Vries	19
	Benno Stein	17
	Maarten de Rijke	17
	Iadh Ounis	16
	Joemon M. Jose	16
	James Allan	15
	Jimmy J. Lin	15
	Norbert Fuhr	15
	Alistair Moffat	14
Publications (4939)	ICTIR	314
	SIGIR	1621
	CIKM	881
	ECIR	625
	TREC	724
	WSDM	197
	AIRS	168
	JCDL	121
	AAAI	185
	CHIIR	103
Company (31)	Twitter	42
	Microsoft	5
	Threshold_Records	5
	Target_Corporation	4
	Xerox	4
	Smart_Communications	3
	Ableton	2
	Associated_Press	2
	Gap_Inc.	2
	Google_Maps	2

## User Study. Comparison to dblp.

14 users performed search tasks in two phases: *P1* simple search tasks, *P2* complex search tasks. Overall, our system *poda* is perceived as good as dblp after *P1*, clearly superior after *P2*.

