A dimly lit museum gallery with large paintings on the wall. A person is walking in the foreground, looking at the art. The lighting is dramatic, highlighting the paintings and the person's silhouette.

Serhii Berehovi, Anselm Nehls


The Da Vinci Code

Auf der Suche nach der *Suche nach dem Code...*

Aufgabe 2


Erstmal wieder zusammensetzen...

...was zusammengehört.



```
CREATE TABLE aol_query (  
  id SERIAL PRIMARY KEY,  
  user_id INTEGER NOT NULL,  
  timestamp TIMESTAMP,  
  text TEXT,  
  page_rank INTEGER,  
  url_id INTEGER REFERENCES url(id)  
);
```

```
1 v create view complete_queries as  
2  
3 select  
4   aol_query.user_id,  
5   aol_query.timestamp,  
6   aol_query.text,  
7   aol_query.page_rank,  
8   url.url  
9 from  
10  aol_query  
11 left join  
12  url on url.id = aol_query.url_id  
13 order by  
14  aol_query.timestamp asc;
```



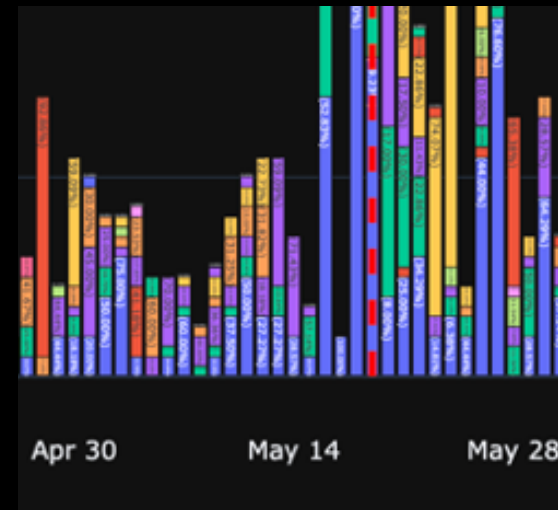
```
CREATE TABLE url (  
  id SERIAL PRIMARY KEY,  
  url TEXT NOT NULL  
);
```

** LEFT JOIN, da nicht alle queries URLs zugeordnet haben*

Schauspieler und Orte

Wie hat sich das Interesse an den Mitwirkenden des Films vor und nach Veröffentlichung verändert?

```
1  select
2      fc.name,
3      count(q.text) as cnt,
4      q.timestamp::date as d
5
6  from
7      aol_query as q
8
9  join
10     film_cast as fc
11  on
12     q.text LIKE '%' || lower(fc.name) || '%'
13
14  group by
15     fc.name,
16     d
17
18  order by
19     fc.name asc,
20     d asc
21  |
```



*Visualisierung in
der nächsten
Aufgabe*

Schauspieler und Orte

Welche Orte aus dem Film wurden besonders häufig gesucht?

```
INSERT INTO location (name) VALUES ('Louvre Museum');
INSERT INTO location (name) VALUES ('Rosslyn Chapel');
INSERT INTO location (name) VALUES ('Temple Church');
INSERT INTO location (name) VALUES ('Château Villette');
INSERT INTO location (name) VALUES ('Lincoln Cathedral');
INSERT INTO location (name) VALUES ('Westminster Abbey');
INSERT INTO location (name) VALUES ('Saint-Sulpice');
INSERT INTO location
INSERT INTO location
INSERT INTO location


select
    loc.name,
    count(*) as location_count
from
    complete_queries as q
join
    location as loc
on q.text like '%' || lower(loc.name) || '%'
group by
    loc.id
order by
    location_count desc
```

	name character varying (255) 🔒	location_count bigint 🔒
1	Temple Church	79
2	Louvre Museum	37
3	Rosslyn Chapel	32
4	Westminster Abbey	31
5	Lincoln Cathedral	6
6	Salisbury Cathedral	3
7	Saint-Sulpice	1

Schauspieler und Orte

Wie hat sich das Interesse an den Orten, die im Film vorkommen, mit der Veröffentlichung verändert?

```
1 select
2     loc.name,
3     count(q.text) as cnt,
4     q.timestamp::date as d
5
6 from
7     complete_queries as q
8
9 join
10    location as loc
11 on
12    q.text LIKE '%' || lower(loc.name) || '%'
13
14 group by
15     q.user_id,
16     loc.name,
17     d
18
19 order by
20     loc.name asc,
21     d asc
22
```



	name character varying (255)	cnt bigint	d date
1	Lincoln Cathedral	5	2006-05-18
2	Lincoln Cathedral	1	2006-05-19
3	Louvre Museum	3	2006-03-05
4	Louvre Museum	5	2006-03-10
5	Louvre Museum	2	2006-03-17
6	Louvre Museum	2	2006-03-19
7	Louvre Museum	2	2006-03-24
8	Louvre Museum	5	2006-03-26
9	Louvre Museum	2	2006-03-29
10	Louvre Museum	2	2006-04-02
11	Louvre Museum	1	2006-04-11
12	Louvre Museum	3	2006-04-27
13	Louvre Museum	2	2006-04-30
14	Louvre Museum	1	2006-05-03
15	Louvre Museum	2	2006-05-04
16	Louvre Museum	1	2006-05-16
17	Louvre Museum	2	2006-05-20
18	Louvre Museum	2	2006-05-30
19	Rosslyn Chapel	3	2006-03-12
20	Rosslyn Chapel	1	2006-03-13
21	Rosslyn Chapel	1	2006-03-20

Ergebnis in dieser Form eher nicht statistisch relevant.

Vermutlich aussagekräftiger, aber aus Zeitgründen nicht mehr umgesetzt, wäre die Verbindung mittels unschärferen *search_terms*

(z.B. "louvre" statt "Louvre Museum")

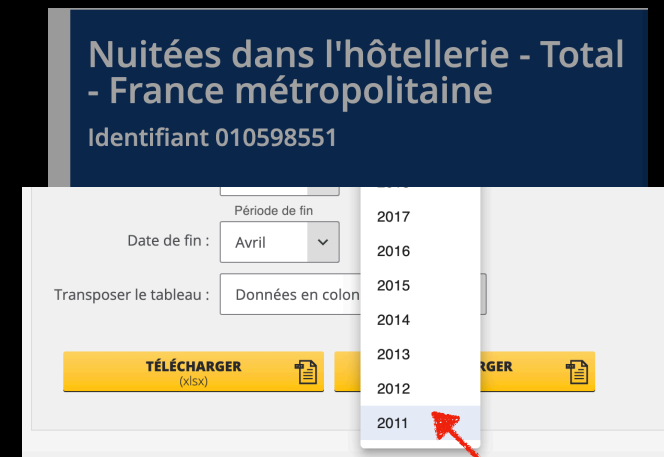
Schauspieler und Orte

Wie hat sich die Veröffentlichung des Films auf den Tourismus an diesen Orten ausgewirkt?

Leider lässt sich diese* Frage nicht so leicht beantworten...



...da entweder die Links zu externen Datensätzen nicht mehr aktiv sind.



...oder aus dem benötigten Zeitraum gar keine Daten (mehr?) vorhanden sind.

* gilt auch für die Frage: "Wie hat sich die Veröffentlichung des Films auf die Einschreibungen in Kunsthochschulen ausgewirkt?"

Filmthemen

Welche Kunstwerke, Symbole oder Orte (=Themen) aus dem Film haben die meisten Suchanfragen hervorgebracht, gruppiert nach Häufigkeit?

```
1 select
2     pt.name,
3     count(*) as query_count
4 from
5     aol_query as aq
6 join
7     plot_theme as pt
8     on aq.text like '%' || lower(pt.name) || '%'
9 group by
10     pt.id
11 order by
12     query_count desc
```



	name character varying (255)	query_count bigint
1	The Last Supper	1357
2	Mona Lisa	694
3	Opus Dei	559
4	Priory of Sion	202
5	The Holy Grail	154
6	Vitruvian Man	118
7	Temple Church	79
8	Madonna of the Rocks	69
9	Louvre Museum	37
10	Rosslyn Chapel	32
11	Sacred Feminine	25
12	Cryptex	24
13	Hieros Gamos	19
14	Jesus and Mary Magdalene	9

Geheimbünde und Verschwörungstheorien

Auf welchen Webseiten wurde am meisten über den Film betreffende Verschwörungstheorien gesprochen bzw. recherchiert?

```
1  select
2    u.url,
3    count(u.url) as url_count
4  from
5    aol_query as aq
6
7  join
8    plot_theme as pt
9    on aq.text like '%' || lower(pt.name) || '%'
10 join
11    url as u
12    on aq.url_id = u.id
13
14 group by
15    u.url
16 order by
17    url_count desc
```



	url text	url_count bigint
1	http://www.artchive.com	228
2	http://www.lisashea.com	194
3	http://en.wikipedia.org	144
4	http://www.opusdei.org	113
5	http://www.odan.org	89
6	http://www.ibiblio.org	68
7	http://www.mond.at	66
8	http://arthistory.about.com	55
9	http://ccat.sas.upenn.edu	49
10	http://www.imdb.com	37
11	http://priory-of-sion.com	34
12	http://www.fiu.edu	34
13	http://milano.arounder.com	31
14	http://www.rickross.com	25
15	http://www.press.uchicago.edu	25
16	http://www.newadvent.org	24
17	http://www.americamagazine.org	23
18	http://www.wfu.edu	22
19	http://www.monalisamania.com	22
20	http://www.aiwaz.net	21
21	http://www.ordotempli.org	20
22	http://www.louvre.fr	19

Geheimbünde und Verschwörungstheorien

Um nicht durch die exakte Schreibweise von Organisationen in Anfragen limitiert zu werden, ordnen wir mögliche Suchbegriffe zu

```
INSERT INTO search_term (text) VALUES
('opus dei'), ('opusdei'),
('priory of sion'), ('sion of priory'), ('prieuré de sion'),
('illuminati'), ('illuminatus'), ('illuminaten');

INSERT INTO search_term_organisation (search_term_id, organisation_id) VALUES
(1, 7), (2, 7), -- Search terms related to Opus Dei
(3, 15), (4, 15), (5, 15), -- Search terms related to Priory of Sion
(6, 2), (7, 2), (8, 2); -- Search terms related to Illuminati
```

```
1 select distinct
2   aq.*
3 from
4   aol_query as aq
5
6 join
7   search_term as st
8   on aq.text like '% ' || st.text || ' %'
9   or aq.text like st.text || ' %'
10  or aq.text like '% ' || st.text
11  or aq.text = st.text
12 join
13   search_term_organisation as sto
14   on st.id = sto.search_term_id
15 join
16   organisation as org
17   on sto.organisation_id = org.id
18
19 where
20   org.in_film = TRUE;
```

	id [PK] integer	user_id integer	timestamp timestamp without time zone	text text	page_rank integer	url_id integer
116	4128461	8715212	2006-05-19 15:39:03	priory of sion	7	580285
117	4128462	1374900	2006-05-28 19:01:27	priory of sion	8	580285
118	4203705	8167270	2006-05-01 14:17:20	opus dei coin	1	2891
119	4261079	6187651	2006-05-19 14:00:40	7 grand masters of the priory of sion	7	2891
120	4275211	7648899	2006-04-30 22:04:02	opusdei	2	443783
121	4275220	2262959	2006-03-22 09:46:12	opus dei	8	443783
122	4275223	190827	2006-03-11 17:47:23	opus dei	6	443783
123	4275227	3443382	2006-05-18 05:18:11	opus dei	6	443783
124	4275228	3289429	2006-05-15 14:03:54	opusdei ministry	4	443783
125	4275229	3289429	2006-05-15 17:24:41	opus dei criticism	4	443783
126	4275230	3610912	2006-04-21 06:44:57	opus dei	6	443783
127	4275236	830699	2006-05-15 19:44:09	opus dei	8	443783

Geheimbünde und Verschwörungstheorien

Auf welche realen Organisationen/Geheimbünde wurde das öffentliche Interesse durch die im Film vorkommenden Organisationen "Opus Dei" und "Priory of Sion" gelenkt?

Anfrage ist so formuliert kaum umsetzbar, daher simpler:

1. Zeige alle User, die zuerst nach einer der Organisationen aus dem Film und später nach einer nicht im Film vorkommenden Geheimorganisation gesucht haben.
2. Zeige alle betreffenden Suchanfragen dieser User, sortiert nach Suchzeitpunkt


Geheimbünde und Verschwörungstheorien


1. Zeige alle User, die zuerst nach einer der Organisationen aus dem Film und später nach einer nicht im Film vorkommenden Geheimorganisation gesucht haben.

```
with UsersWhoSearchedForOrgsInFilm as (  
  select  
    aq.user_id,  
    min(aq.timestamp) as first_search  
  from  
    aol_query as aq  
  join  
    search_term as st  
    on aq.text like '% ' || st.text || ' %'  
    or aq.text like st.text || ' %'  
    or aq.text like '% ' || st.text  
    or aq.text = st.text  
  join search_term_organisation as sto  
    on st.id = sto.search_term_id  
  join organisation as org  
    on sto.organisation_id = org.id  
  where  
    org.in_film = TRUE  
  group by  
    aq.user_id  
)
```

```
UsersWhoSearchedForOrgsNotInFilm as (  
  select  
    aq.user_id,  
    min(aq.timestamp) as first_search  
  from  
    aol_query as aq  
  join  
    search_term as st on aq.text like '% ' || st.text || ' %'  
    or aq.text like st.text || ' %'  
    or aq.text like '% ' || st.text  
    or aq.text = st.text  
  join  
    search_term_organisation as sto  
    on st.id = sto.search_term_id  
  join organisation org  
    on sto.organisation_id = org.id  
  where  
    org.in_film = FALSE  
  group by  
    aq.user_id  
)
```

```
select  
  oif.user_id  
from  
  UsersWhoSearchedForOrgsInFilm as oif  
join  
  UsersWhoSearchedForOrgsNotInFilm as onif  
  on oif.user_id = onif.user_id  
where  
  oif.first_search < onif.first_search;
```



	user_id 
1	20753228
2	21299479
3	694246
4	1609874
5	1754281

Geheimbünde und Verschwörungstheorien

2. Zeige alle betreffenden Suchanfragen dieser User, sortiert nach Zeitpunkt

```
1 select
2   aq.*
3 from
4   aol_query as aq
5
6 join
7   search_term as st
8   on aq.text like '%' || st.text || '%'
9 join
10  search_term_organisation as sto
11  on st.id = sto.search_term_id
12 join
13  organisation as org
14  on sto.organisation_id = org.id
15
16 where
17   aq.user_id in (20753228, 21299479, 694246, 1609874, 1754281)
18
19 order by
20   aq.user_id, aq.timestamp;
```

*Ja, könnte man auch als
weitere subquery einsetzen...*


	user_id integer
1	20753228
2	21299479
3	694246
4	1609874
5	1754281

	id [PK] integer	user_id integer	timestamp timestamp without time zone	text text
1	14200466	694246	2006-05-19 23:04:45	opus dei
2	7146290	694246	2006-05-19 23:15:46	secta illuminati
3	1033191	1609874	2006-04-18 00:40:11	opus dei
4	14200572	1609874	2006-04-18 00:40:11	opus dei
5	12835501	1609874	2006-04-18 00:40:11	opus dei
6	30518525	1609874	2006-04-18 00:58:05	opus dei
7	11011830	1609874	2006-05-22 02:49:14	illuminati
8	1033175	1609874	2006-05-22 02:49:14	illuminati
9	10739192	1609874	2006-05-22 02:49:14	illuminati
10	8922916	1609874	2006-05-22 04:53:51	illuminati
11	14062013	1754281	2006-03-05 06:18:46	opus dei
12	19844542	1754281	2006-03-05 06:20:25	opus dei
13	6838345	1754281	2006-03-05 06:21:26	illuminati
14	14200602	20753228	2006-04-19 19:08:39	opus dei
15	8674453	20753228	2006-04-26 10:13:41	priory of sion
16	6838346	20753228	2006-04-28 11:12:24	illuminati
17	4275306	21299479	2006-04-23 19:09:10	opus dei
18	34743277	21299479	2006-04-23 19:28:09	illuminati
19	34743278	21299479	2006-04-23 19:33:35	illuminati

Geheimbünde und Verschwörungstheorien

Wieviele Benutzer haben sich im untersuchten Zeitraum darüber informiert, wie sie einem Geheimbund beitreten können?

```
1 select
2   q.*
3 from
4   complete_queries as q
5 join
6   search_term as st
7   on q.text like '% ' || st.text || ' %'
8   or q.text like st.text || ' %'
9   or q.text like '% ' || st.text
10  or q.text = st.text
11 join
12   search_term_organisation as sto
13   on st.id = sto.search_term_id
14 join
15   organisation as org
16   on sto.organisation_id = org.id
17
18 where
19   q.text like '% ' || 'join' || ' %'
20   or q.text like '% ' || 'member' || ' %'
```



	user_id integer	timestamp timestamp without time zone	text text	page_rank integer	url text
1	9925869	2006-03-28 13:45:59	illuminati members	5	http://www.illuminati-news.com
2	3726202	2006-05-19 22:13:21	opus dei members wear a pin	4	http://lfpress.ca

Ergebnis: keine :(

Geheimbünde und Verschwörungstheorien

Wie groß war der durchschnittliche Anteil an auf Themen aus dem Film bezogenen Suchanfragen im Verhältnis zu anderen Suchanfragen bei Benutzern, die mindestens einmal nach einem auf den Film bezogenen Thema gesucht haben?

```
with FilmRelatedQueries as (  
  select  
    aq.user_id, aq.id  
  from  
    aol_query as aq  
  join  
    plot_theme as pt  
    on aq.text like '%' || lower(pt.name) || '%'  
)  
  
UsersWithFilmRelatedQueries as (  
  select distinct  
    user_id  
  from  
    FilmRelatedQueries  
)
```

```
TotalQueriesCount as (  
  select  
    user_id,  
    count(*) as total_queries  
  from  
    aol_query  
  where  
    user_id in  
      (select user_id from UsersWithFilmRelatedQueries)  
  group by  
    user_id  
)  
  
FilmRelatedQueriesCount as (  
  select  
    user_id,  
    count(*) as film_queries  
  from  
    FilmRelatedQueries  
  group by  
    user_id  
)
```

```
select  
  related_cnt.user_id,  
  cast(related_cnt.film_queries as FLOAT) / total_count.total_queries as proportion  
from  
  TotalQueriesCount as total_count  
join  
  FilmRelatedQueriesCount as related_cnt  
on total_count.user_id = related_cnt.user_id;
```

	user_id integer	proportion double precision
1	14756	0.02
2	17138	0.018518518518518517
3	19166	0.004651162790697674
4	23292	0.006968641114982578
5	28707	0.00522875816993464
6	28829	0.04716981132075472
7	31363	0.025210084033613446
8	56833	0.007633587786259542
9	63119	0.006060606060606061
10	67137	0.012875536480686695
11	71845	6.08381347743621e-05
12	89214	0.005
13	91008	0.08064516129032258
14	99529	0.1935483870967742
15	120744	0.02099737532808399
16	123703	0.0008022462896109105
17	132661	0.006578947368421052
18	135020	0.05
19	144956	0.018867924528301886
20	149317	0.011049723756906077
21	151300	0.09090909090909091
22	190827	0.004351610095735422
23	200597	0.24836601307189543
Total rows: 1000 of 1035 Query complete		

Teil 1: wir suchen die Verhältnisse von Filmthemen zu anderen Themen aller User, die nach Filmthemen gesucht haben

Geheimbünde und Verschwörungstheorien

Wie groß war der durchschnittliche Anteil an auf Themen aus dem Film bezogenen Suchanfragen im Verhältnis zu anderen Suchanfragen bei Benutzern, die mindestens einmal nach einem auf den Film bezogenen Thema gesucht haben?

```
select * from (  
  select  
    related_cnt.user_id,  
    cast(related_cnt.film_queries as FLOAT) / total_count.total_queries as proportion  
  from  
    TotalQueriesCount as total_count  
  join  
    FilmRelatedQueriesCount as related_cnt  
    on total_count.user_id = related_cnt.user_id  
  ) as results  
where  
  proportion > '0.1'  
order by  
  proportion desc;
```

Teil 2: wir suchen die User, die besonders aktiv nach Filmthemen gesucht haben

	user_id integer	proportion double precision
1	24597382	1
2	21402603	1
3	9726136	0.6153846153846154
4	22183599	0.5
5	20951883	0.5
6	3938103	0.4827586206896552
7	4772460	0.3953488372093023
8	2648974	0.3333333333333333
9	14965140	0.3333333333333333
10	20057853	0.3333333333333333
11	23423851	0.3275862068965517
12	18512000	0.3076923076923077
13	20118302	0.3
14	20959062	0.2857142857142857
15	943701	0.2857142857142857
16	13491856	0.2727272727272727
17	22649060	0.2631578947368421
18	8793094	0.26229508196721312


Die beiden haben nur einmal AOL benutzt...

Geheimbünde und Verschwörungstheorien

Wie groß war der durchschnittliche Anteil an auf Themen aus dem Film bezogenen Suchanfragen im Verhältnis zu anderen Suchanfragen bei Benutzern, die mindestens einmal nach einem auf den Film bezogenen Thema gesucht haben?

```
select
  avg(cast(related_cnt.film_queries as FLOAT) / total_count.total_queries)
  as proportion
from
  TotalQueriesCount as total_count
join
  FilmRelatedQueriesCount as related_cnt
on total_count.user_id = related_cnt.user_id;
```

Teil 3: wir berechnen den
Gesamtdurchschnitt



proportion	🔒
double precision	
0.03243138175277434	