

## Learning Layers

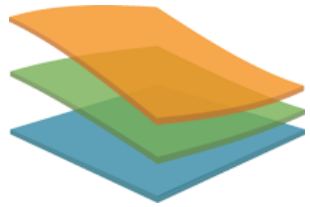
Scaling up Technologies for Informal Learning in SME Clusters

# Towards a Scalable Social Recommender Engine for Online Marketplaces

## The Case of Apache Solr

Emanuel Lacic, **Dominik Kowald**, Denis Parra, Christoph Trattner





# Many thanks to



**Emanuel Lacic**

[elacic@know-center.at](mailto:elacic@know-center.at)

Graz University of Technology  
Austria



**Denis Parra**

[dparra@ing.puc.cl](mailto:dparra@ing.puc.cl)

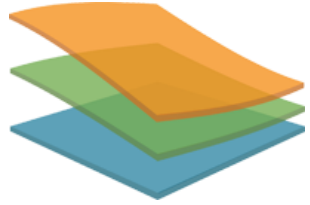
Pontificia Universidad  
Catolica  
Chile



**Christoph Trattner**

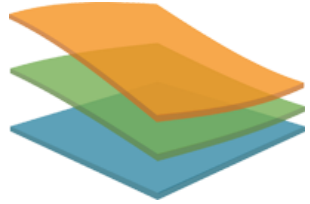
[ctrattner@know-center.at](mailto:ctrattner@know-center.at)

Know-Center Graz  
Austria



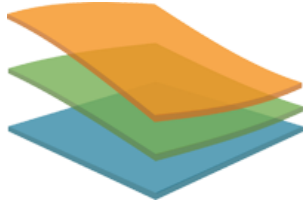
# What will this talk be about?

- (Real-time) product recommendations for online marketplaces
- Scalability of recommender systems
- Utilizing social network data for the recommendations of products to people

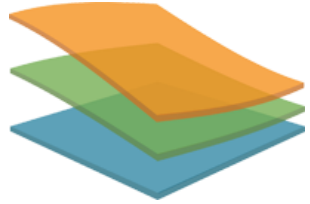


# How did this work start?

- Joint project with the Austrian start-up Blanc-Noir
- Personalized product recommender for online marketplaces based on
  - Actions in the marketplaces (e.g., Ebay, Amazon)
  - Product information
  - Social network data (e.g., Facebook, G+)
  - Filter criteria
- Provided at (near) **real-time!**
  - ... especially if there is a **lot of data**
  - ... together with many **data updates**

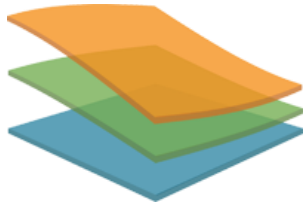


So now, how we have solved that issue?

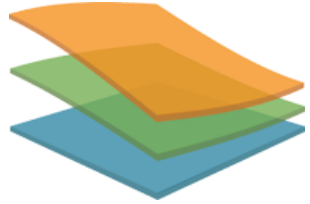


# What's available out there?

- Frameworks/approaches for scalable recommendations
  - Distributed data processing
    - Apache Hadoop / Mahout (map/reduce paradigm)
  - Relational databases
    - MySQL, PostgreSQL (e.g., RecDB project)
  - Collaborative Filtering improvements
    - Matrix factorization
- **Lack of a framework / approach that combines all things we need**



# Apache Solr



# Why Solr?

- „High-performance, full-featured text search engine library“

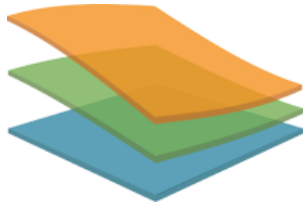
... *but more precise ...*

- „High-performance, fully-featured token **matching** and **scoring** library“ [Grainger, 2012]

... *which provides ....*

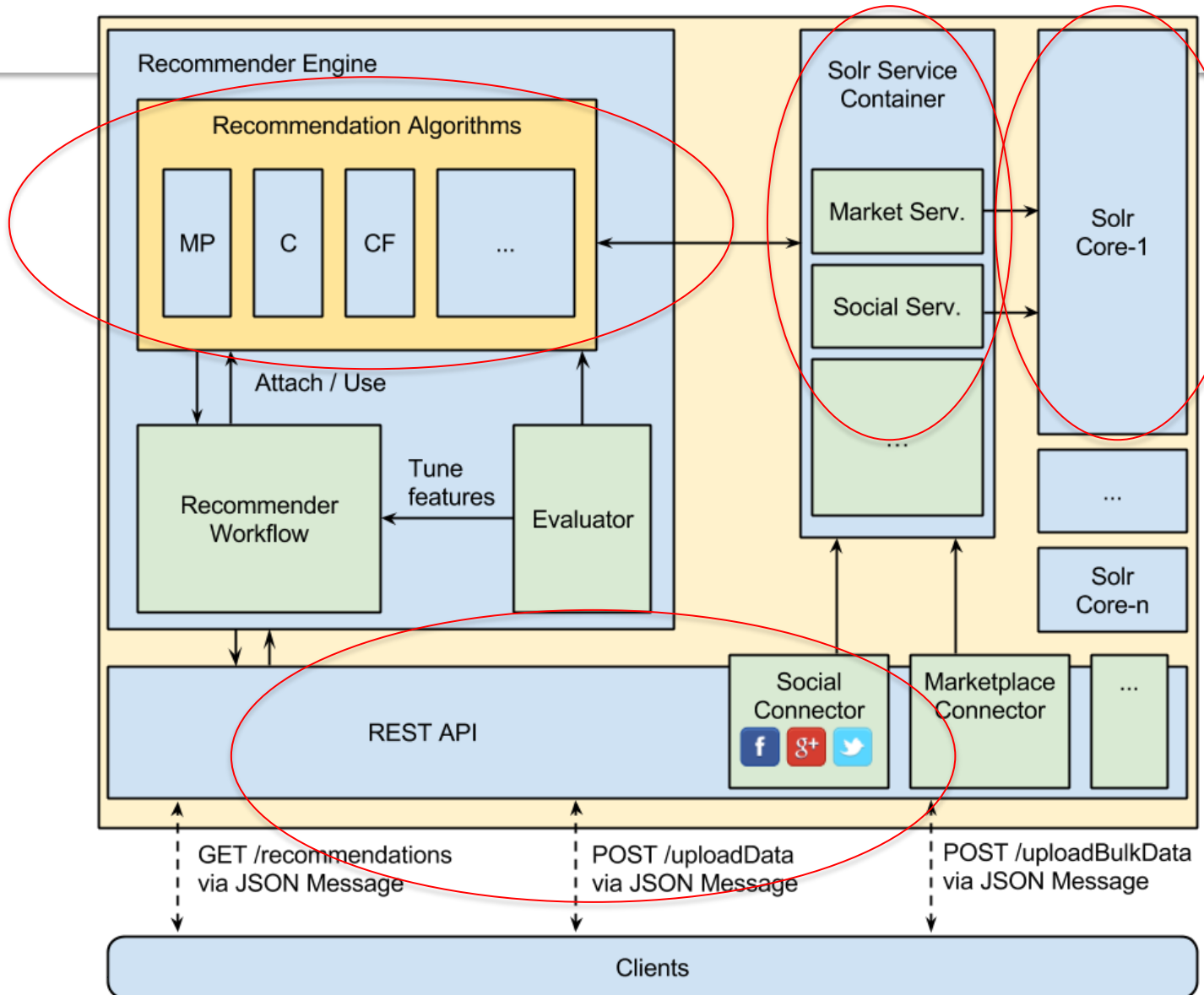
- full-text searches (content-based)
  - powerful queries (e.g., MoreLikeThis or Facets)
  - (near) real-time data updates (no pre/re-calculations)
  - easy schema updates (social data integration)
- Established **open-source software** (Apache license) with big community

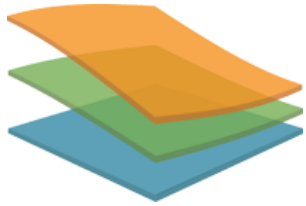




# Our framework

<https://github.com/learning-layers/SocRec>





# How does the thing perform?

- Dataset of virtual world SecondLife
  - Marketplace and social data

Language Home | Help What is Second Life? Join Now. It's Free! Sign In

SECOND LIFE MARKETPLACE

Items Merchants/Stores Cart (0)

Search in: Animals Show maturity levels: General Keywords: Search

36535 matching items found. You may be able to see more items after you log in or join Second Life.

Items per page: 12 Sort by: Relevance

Featured Items

- Bloodhound II**  
VKC Bloodhound II Delivery Kennel - Artificially Intel...  
★★★★★ Reviews (2) L\$4,800  
Virtual Kennel Club (VKC) by Enrico Genosse
- Deer Family**  
Adjustable moving area  
Walking at random  
Blinking eyes  
Grazing  
Copyable  
★★★★★ Reviews (6) L\$350  
Tomato Park-Animals and Pets
- [SYM] Mesh Nylon Harness for WereHouse Dire Wolf**  
★★★★★ Reviews (0) L\$150  
StaloneWolf

Price

L\$0 - L\$10	3547
L\$11 - L\$100	14311
L\$101 - L\$500	13729
L\$501 - L\$1,000	3305
L\$1,001 - L\$5,000	1634
Over L\$5,000	109

Prim count:

Sylvia Tamalyn  
sylvia.tamalyn

Add Friend Follow

Message

Feed About Picks Groups

“Are you wearing your fur coat again, Tem? :p” on dresden.ceriano's snapshot. about 4 hours ago

sylvia.tamalyn shared a snapshot.

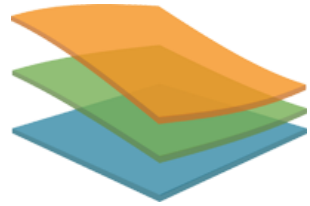
Might as well play with WL while I am here :D  
Shermerville NW Visit location

about 15 hours ago • Love • Comment

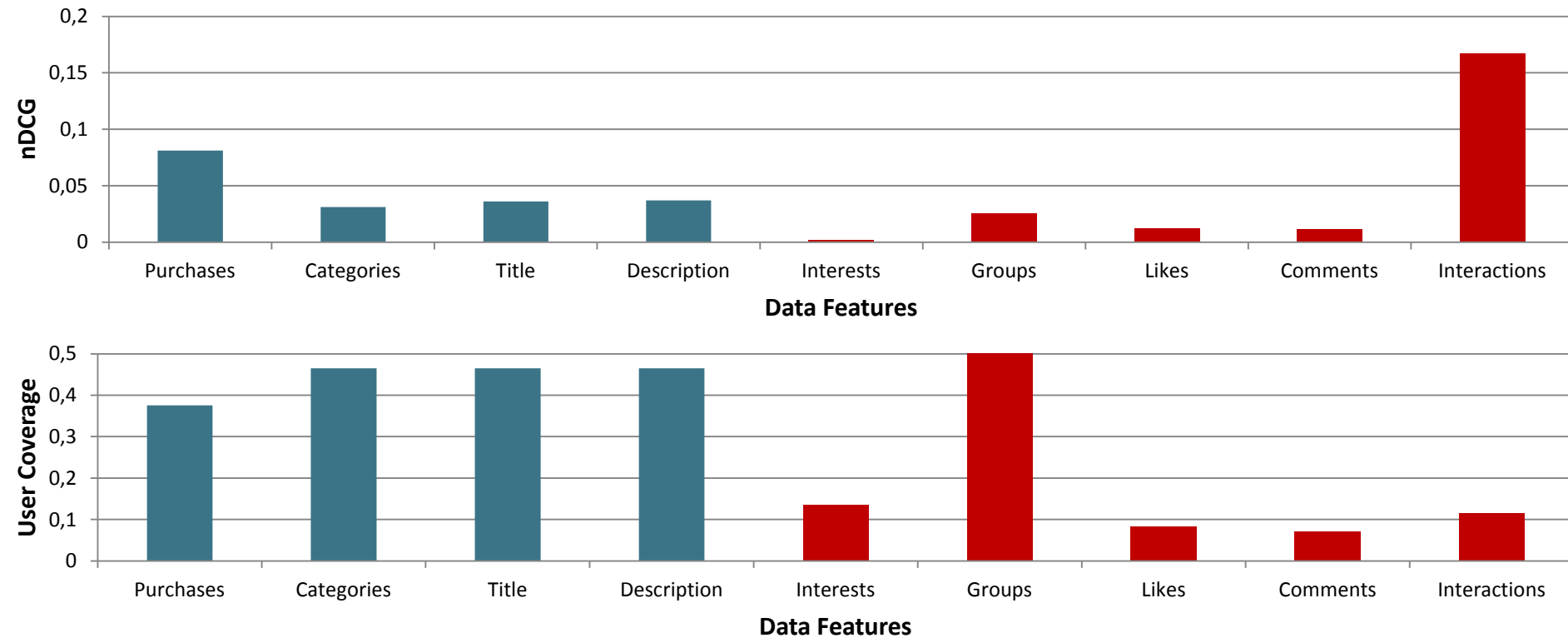
empressazy love the imagery. man i am still workin with WL.  
about 11 hours ago • Love

sylvia.tamalyn Thank you :)  
about 11 hours ago • Love

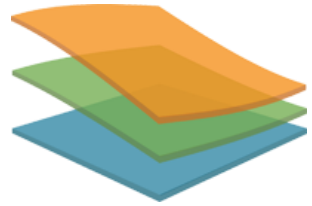
Write a comment...



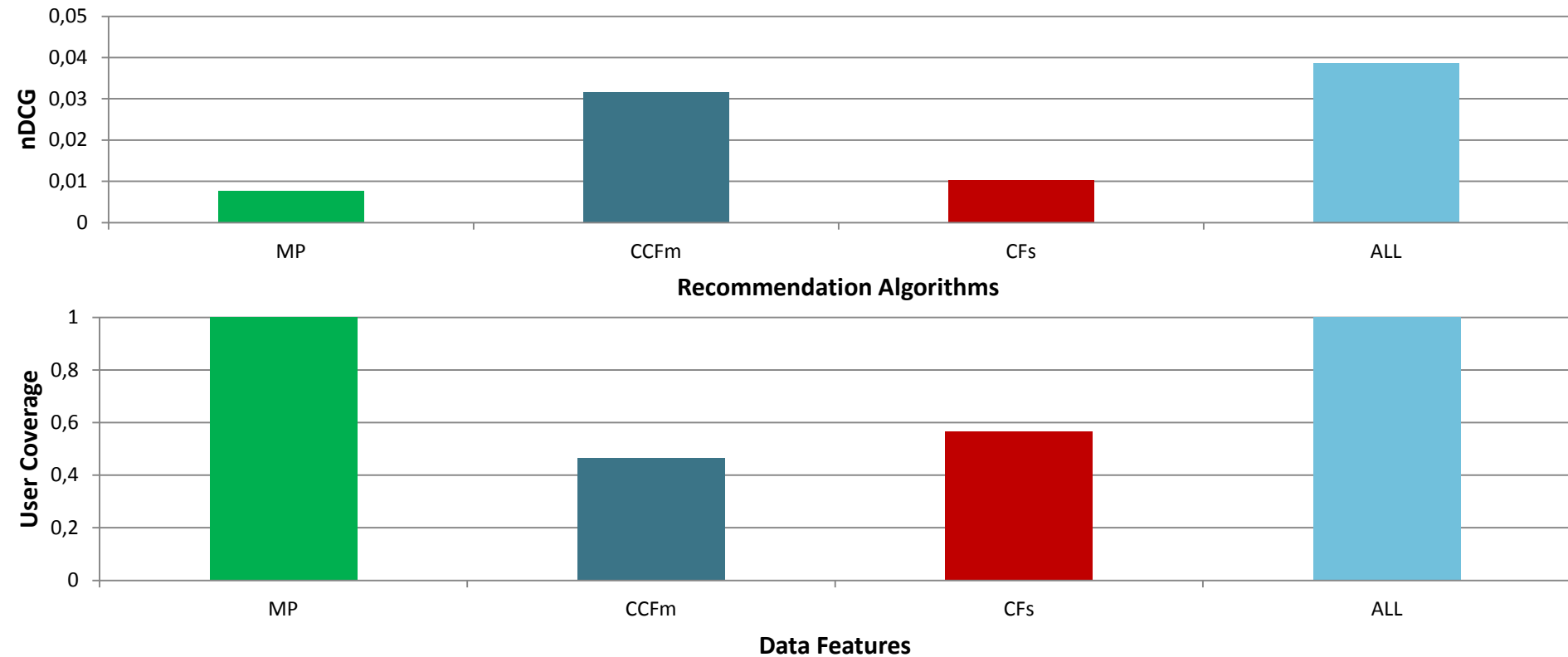
# What's about the marketplace and social data features?



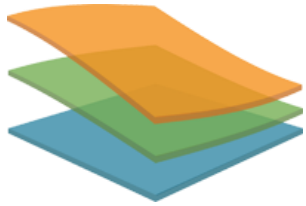
- Both types of data are important for the recommender quality and user coverage



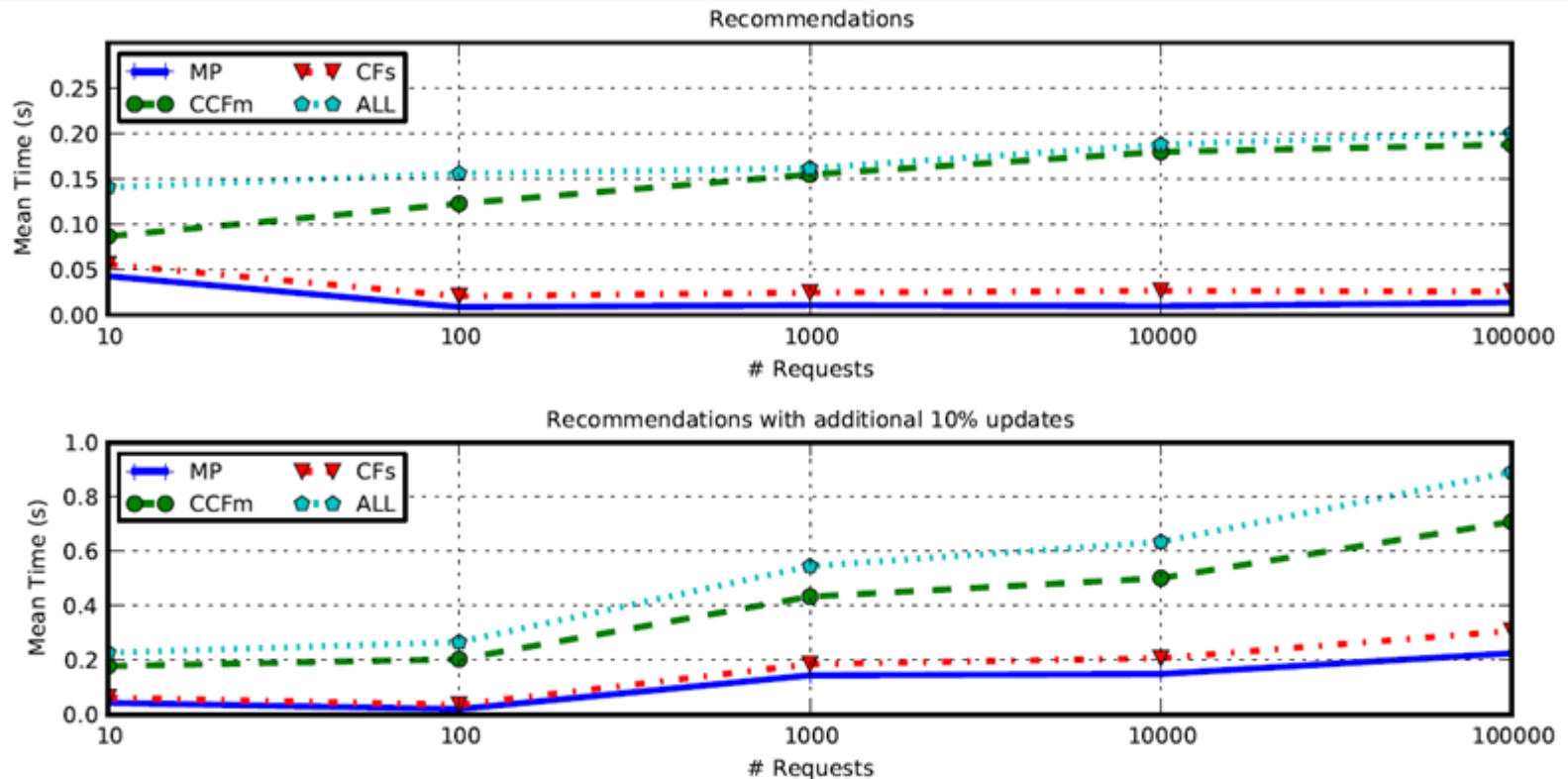
# What's about the hybrids?



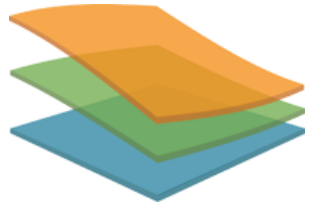
- The hybrid approach provides a good trade-off of recommender quality and user coverage



# What's about the scalability?

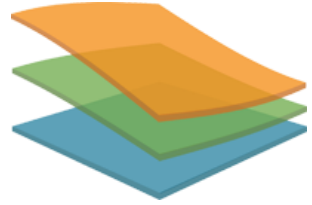


- Recommendations can be provided in (near) real-time in both cases (with and without data update)



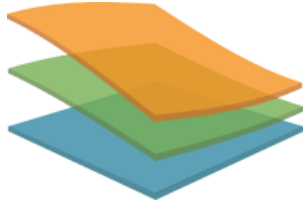
# What we have shown!

- Apache Solr is more than a search engine!
- Actually it is a great framework to implement a **scalable recommender engine for online marketplaces**
  - Near real-time recommendations through build-in query-functions
  - Near real-time **data updates**
  - Easy integration of **social data**  
+ a high-performance **full-text search engine** for free!
- Evaluation on dataset gathered from **SecondLife**
  - Different **marketplace** and **social** data features are important
  - **Hybrid approaches** produce more robust recommendations
  - **It scales!**



# What do we want to do in the future?

- Online study together with BlancNoir with “real” data
- Impact of geo-spatial data
- Impact of temporal data (see WebScience track)
- Comparative study with other backend solutions (e.g., ElasticSearch)



# Thank you for your attention!

**Code and framework:**

<https://github.com/learning-layers/SocRec>

**Questions?**

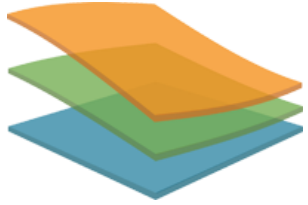
**Dominik Kowald**

[dkowald@know-center.at](mailto:dkowald@know-center.at)

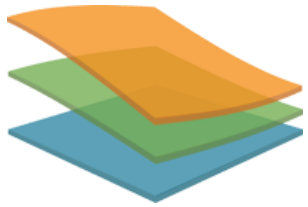
Know-Center

Graz University of Technology (Austria)





# Backup



# Short hands-on session

- Collaborative Filtering

---

```
// Find similar users based on purchased items using
  Solr's facet queries
/select?q=id:("some_product_1")+OR+id:("some_product_2") &
  facet=true&facet.field=my_users_field
// Find items purchased by those similar users that are
  new to the target user
/select?q=my_users_field:("user_1"^5+OR+"user_2"^3) &
  fq:-id:("some_product_1")+OR+-id:("some_product_2")
```

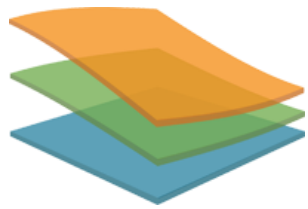
---

- Content-Based

---

```
/select?q=id:("some_product_id") &mlt=true &
  mlt.fl=description
```

---



# SecondLife dataset

---

## Marketplace (Market)

---

Number of users	72,822
Number of purchases	265,274
Mean number of purchases per user	3.64
Number of products	122,360
Mean number of purchases per products	2.17

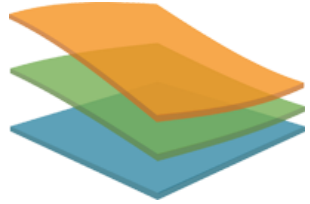
---

## Online Social Network (Social)

---

Number of users	64,500
Number of likes	1,492,028
Number of comments	347,755
Mean number of likes per user	14.91
Mean number of comments per user	3.47
Number of groups	260,137
Mean number of groups per user	8.91
Number of interests	88,371
Mean number of interests per user	1.57

---



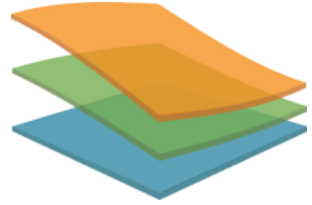
# How to Use the Engine?

- Implement and run a new recommender

---

```
// Implement the recommender strategy
public interface RecommendStrategy {
    public RecommendResponse recommend(RecommendQuery q,
        Integer maxResults, SolrServer SolrServer);
}
// Run the new recommender strategy
RecommendStrategy strategyToUse = new MyStrategyImpl();
Filter filter = new ContentFilter(); // optional
RecommendationService.getRecommendations("some_user",
    "some_product", 10, filter, strategyToUse);
```

---



# Recommendation Algorithms implemented in the Engine

- MostPopular (MP)
  - Recommends for any user the most purchased items
- Collaborative Filtering (CF)
  - Find similar users ( $k$  nearest neighbors) and recommend novel items of those users [Schafer et al., 2007]
  - In Solr: select queries and facet counts
- Content-Based (C)
  - Analyze item meta-data to find similar items [Pazzani et al., 2007]
  - In Solr: *MoreLikeThis* function
- Hybrid (CCF)
  - Combine different algorithms to overcome their individual limitations [Burke et al., 2002]
  - Each algorithm can be weighted / tuned according to its performance