

n Characters in Search of an Author

Improving Author Name Indexing & Searching for the ADS 2.0





Waterwheels, Ace



Wäterwheels, Ace

Häcker, J



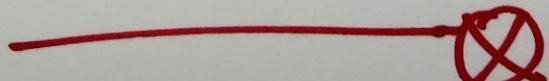
Hacker, J

Zoia Horn



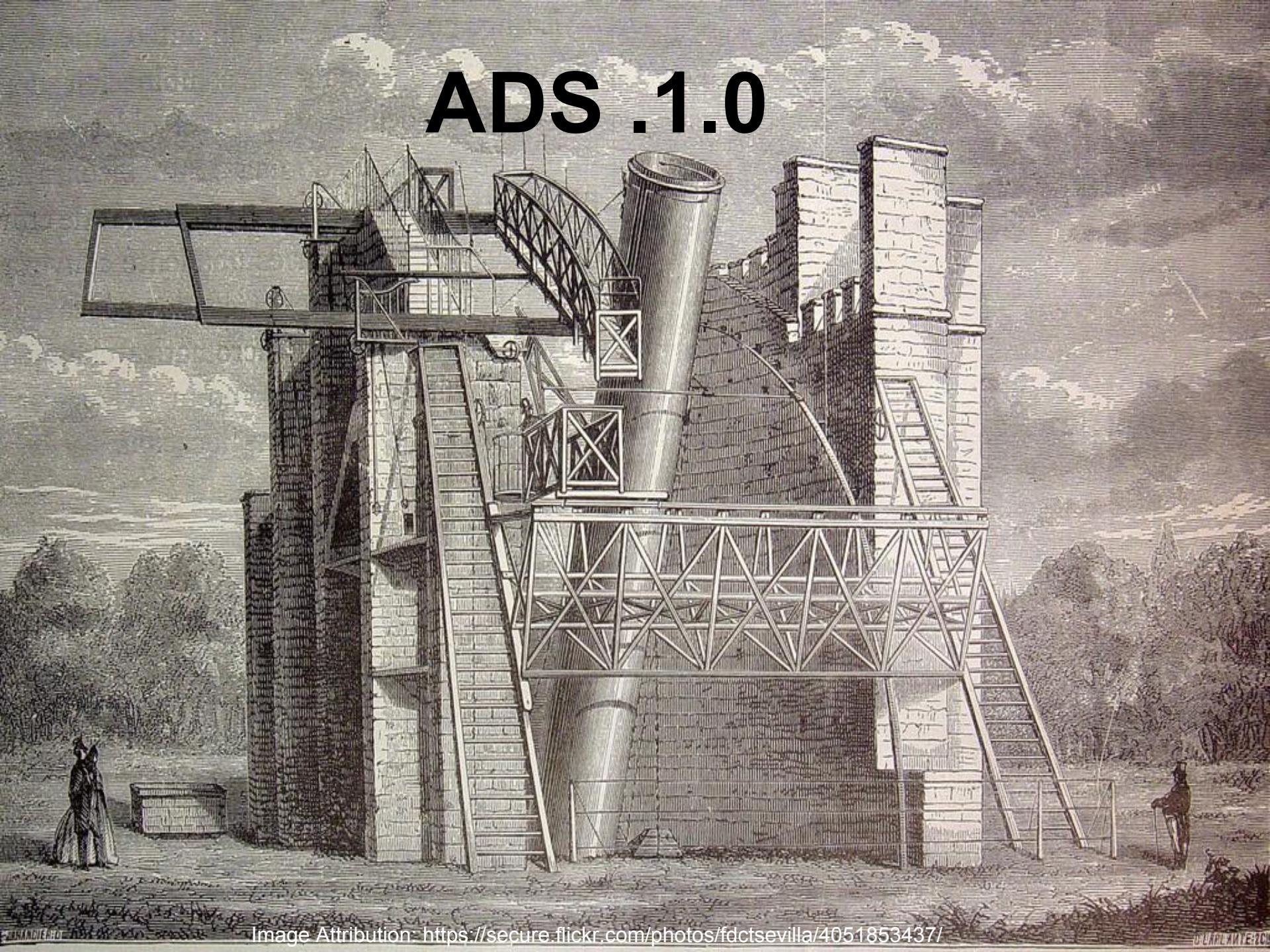
Horn, Z

Billz, Jimmy



Billz, James

ADS .1.0





ADS 2.0

Requirements Gathering

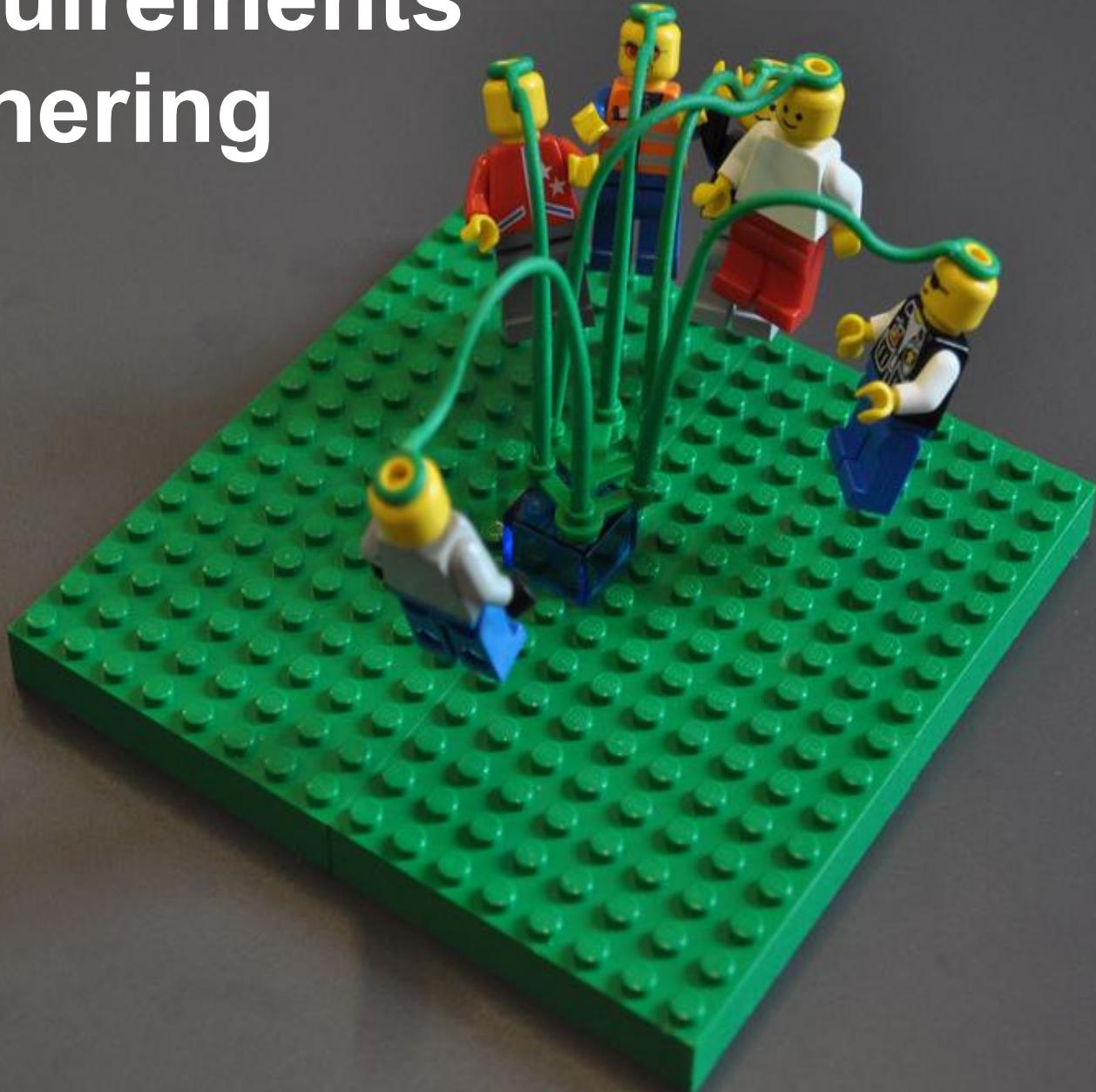


Image Attribution: <https://secure.flickr.com/photos/webatelier/5929855252/>

Req. #1



Hacker, J



Häcker, J

Req. #2



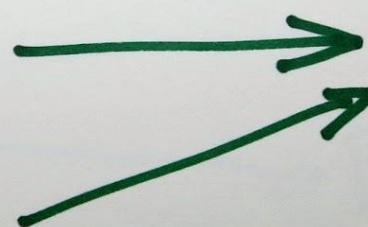
Hørn, Zoia → Horn, Zoia

Req. #3



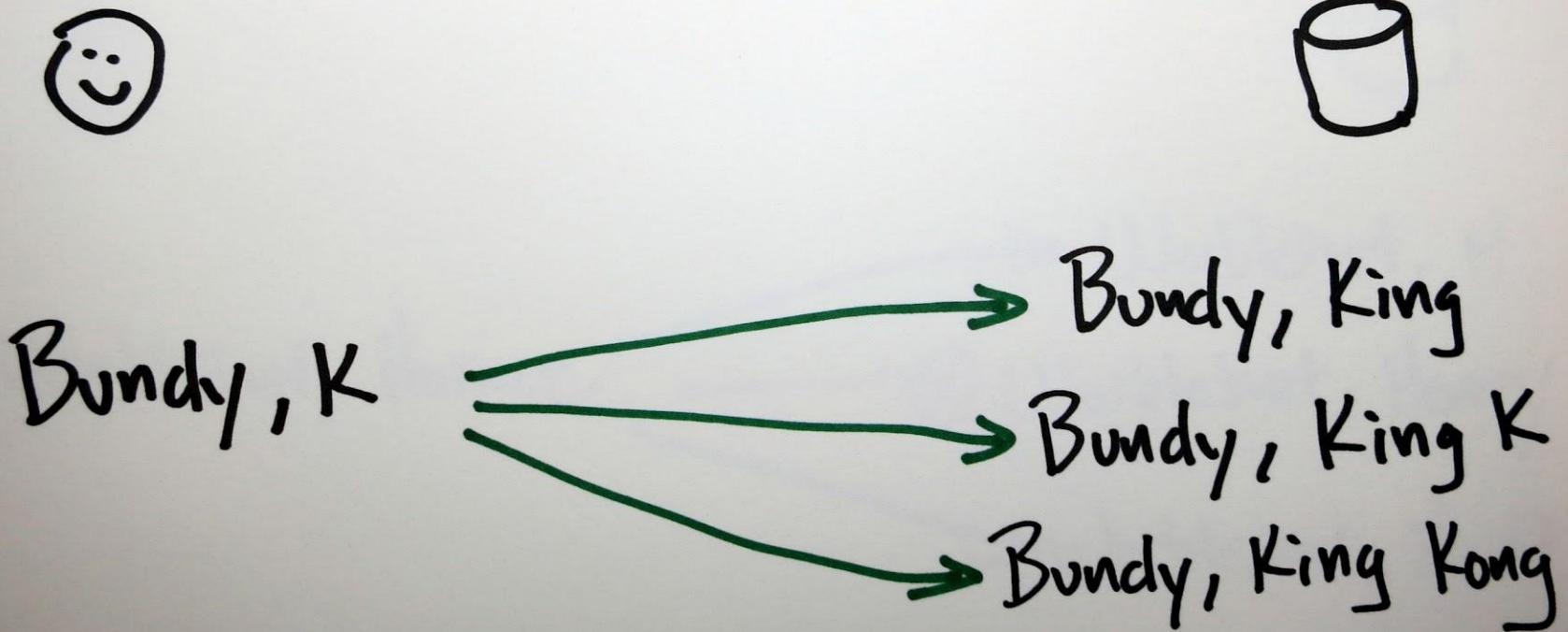
Waterwheels, Ace

Waterwheels, Andy

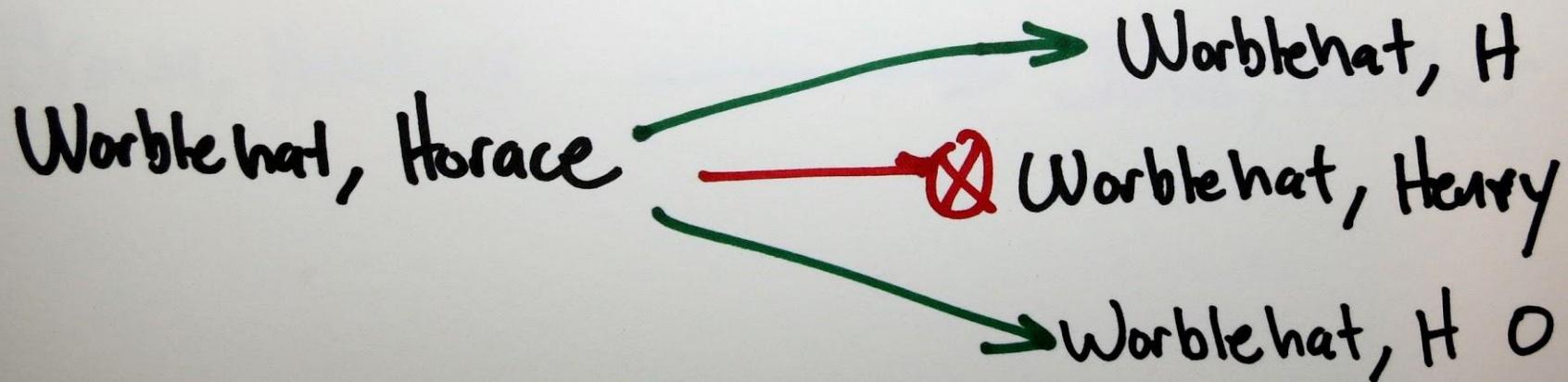


Waterwheels, A

Req. #4



Req. #5



Req. #6



Brown, Heather



Jones, Heather

Req. #7



Exact Match



Malmsteen, Yngwie →⊗ Malmsteen, Yngwie
→⊗ Malmsteen, YJ

Problem Solved!



Normalization

John-Hacker → Hacker, John

Yogi, M.M. → Yogi, MM

Hørn, Zoia; → Hørn, Zoia

Brown-Jones, H → Brown Jones, H

Name Part Variations

Yogi, Maharishi M → [Yogi, Maharishi M*
Yogi, M M*
Yogi, Maharishi
Yogi, M
Yogi,]

Transliteration (i)

Hübbard, M → [Hubbard, M
Huebbard, M]

č → [c, ch]

∅ → [o, oe]

...

Transliteration (2)

Hubbard, M

???

[Hubbard, M
Huebbard, M]

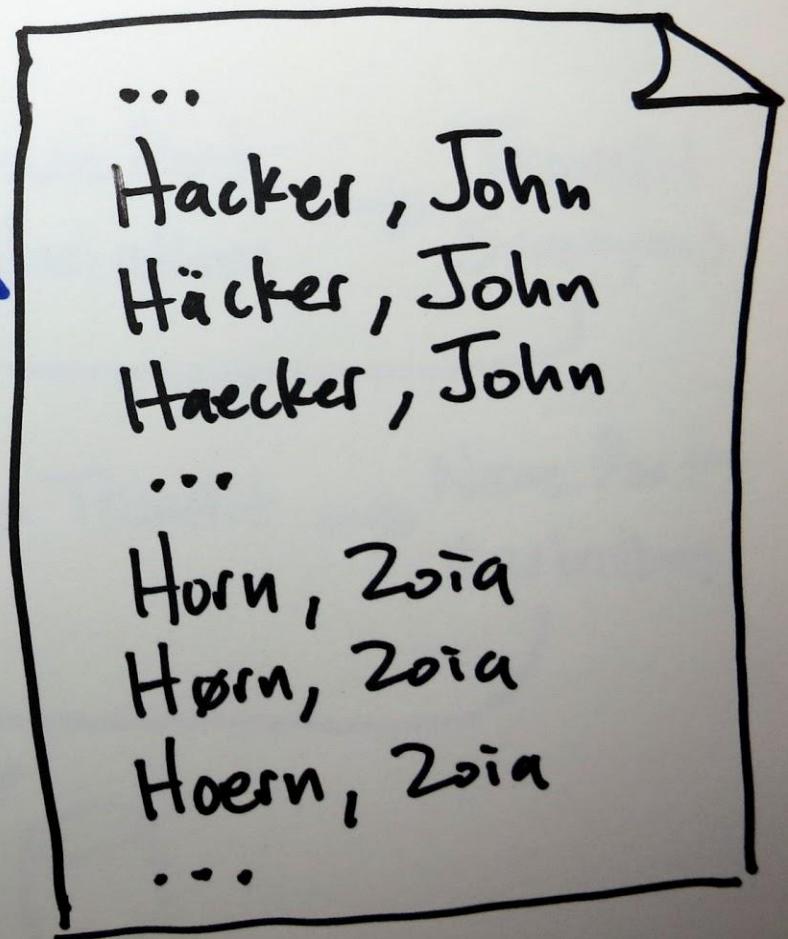
@takify

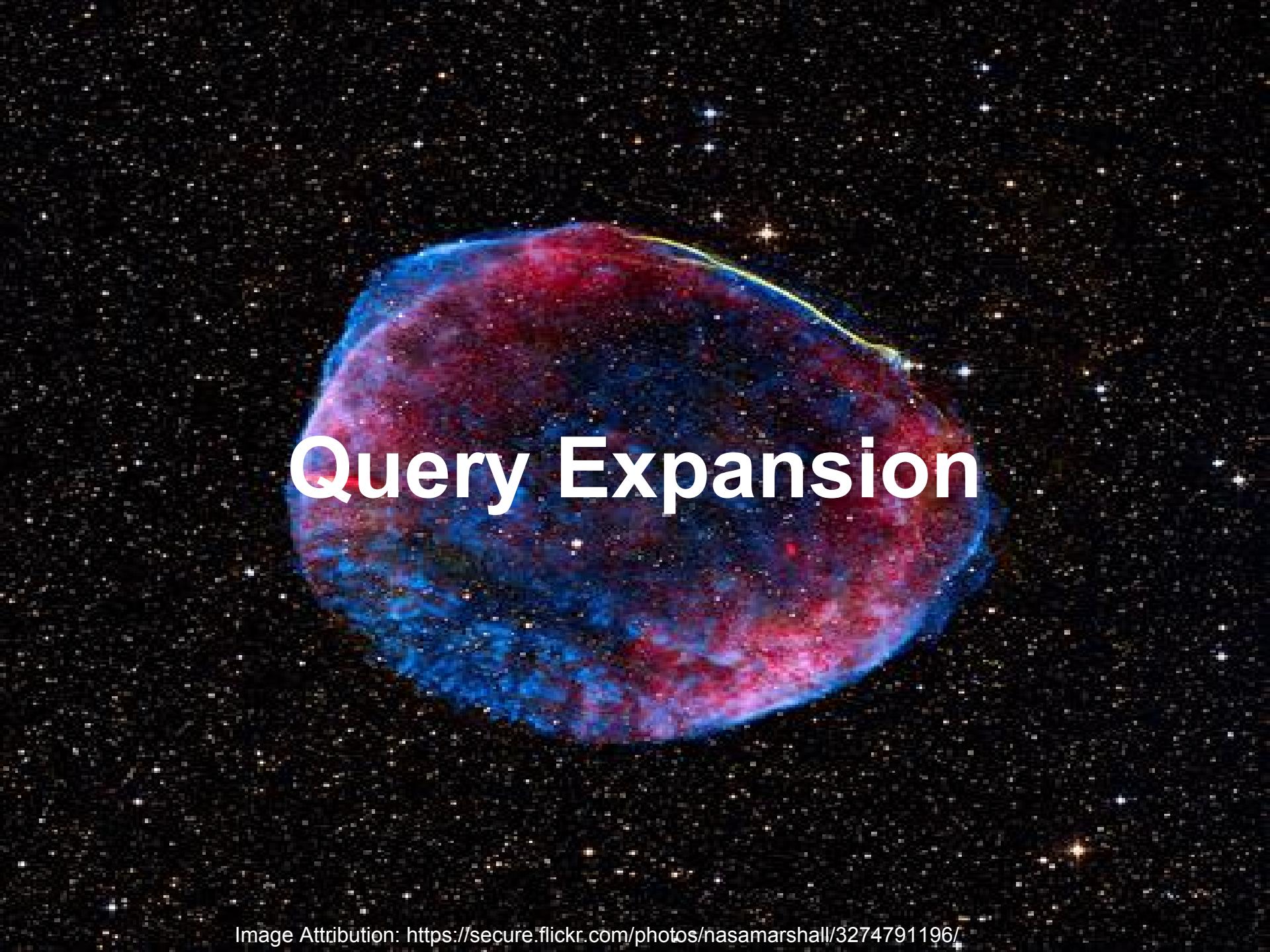
Index Introspection



Häcker, John

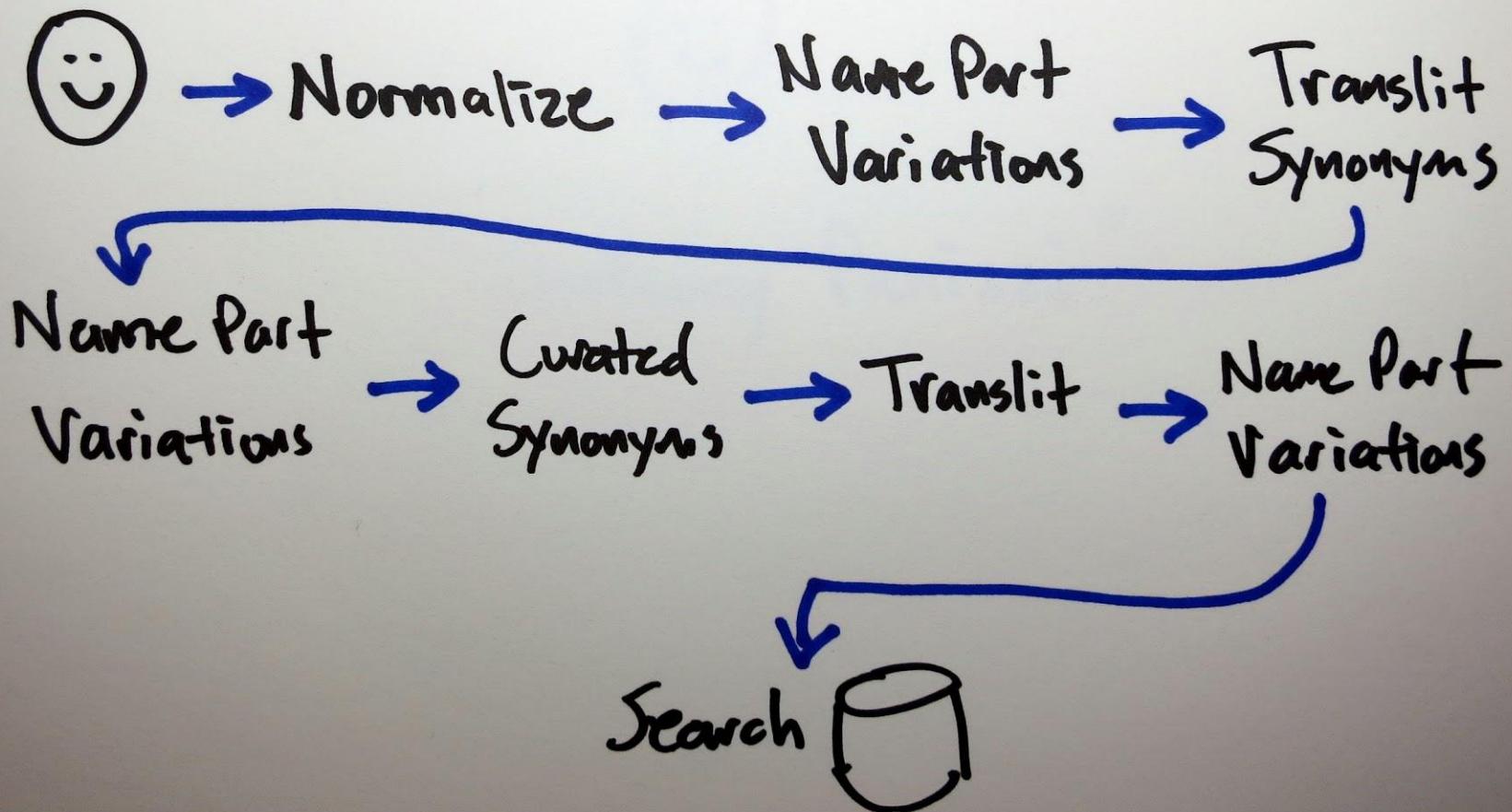
Hørn, Zoia





Query Expansion

Walkthrough



User Input



"Anthony Panizzi"

Normalize



Panizzi, Anthony

Name Part Variations



Panizzi, Anthony

+ Panizzi, Anthony *

+ Panizzi, A

+ Panizzi, A *

+ Panizzi,

Transliterated Synonyms



Panizzi, Anthony
+ Paenizzi, Anthony
+ Pänizzi, Anthony
Panizzi, Anthony *

Panizzi, A
Panizzi, A *

Panizzi,

...
Panizzi, Anthony
Paenizzi, Anthony
Pänizzi, Anthony
...
Krahs, Richard
Kraehs, Richard
Kräbs, Richard
...

Name Part Variations



- | | |
|-----------------------|-----------------|
| Panizzi, Anthony | + Pänizzi, A |
| Paezizzi, Anthony | Panizzi, A * |
| Pünizzi, Anthony | + Paenizzi, A * |
| Panizzi, Anthony * | + Pänizzi, A * |
| + Paezizzi, Anthony * | Panizzi, |
| + Pänizzi, Anthony * | + Paenizzi, |
| Panizzi, A | + Pänizzi, |
| + Paenizzi, A | |

Curated Synonyms



Panizzi, Anthony
Paeizzi, Anthony
Pänizzi, Anthony
...
+ Pänizzi, Tony

...
Kräbs, Richard
Kräbs, Dickie L
...
Pänizzi, Anthony
Pänizzi, Tony
...



Transliteration



Panizzi, Anthony

Pænizzi, Anthony

Pänizzi, Anthony

...

+ Panizzi, Tony

+ Pænizzi, Tony

Pänizzi, Tony

Name Part Variations



Panizzi, Anthony

Paenizzi, Anthony

Pänizzi, Anthony

...

Panizzi, Tony

Paenizzi, Tony

Pänizzi, Tony

+ Panizzi, Tony *

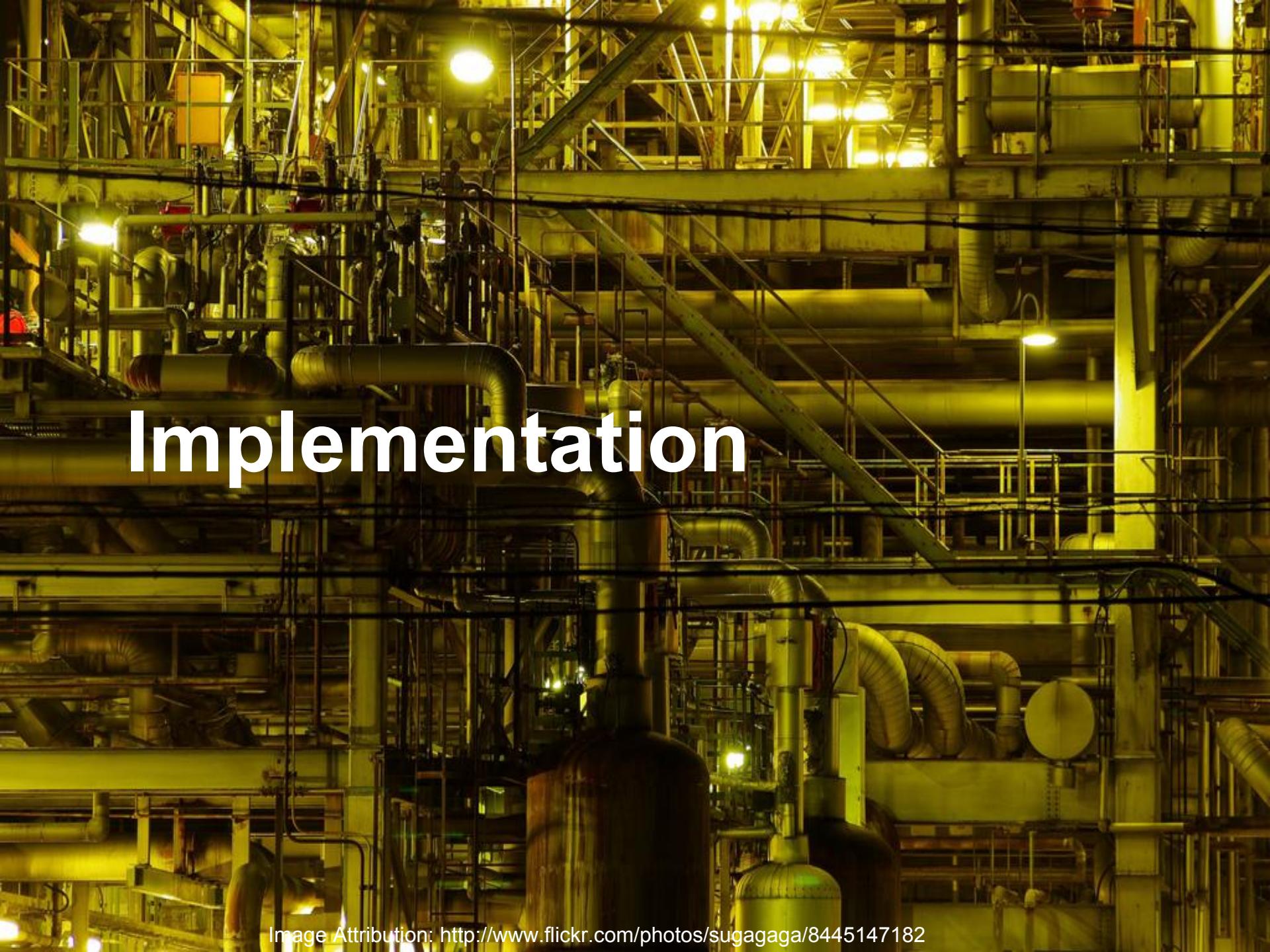
+ Paenizzi, Tony *

+ Pänizzi, Tony *

+ Panizzi, T *

+ Paenizzi, T *

+ Pänizzi, T *



Implementation

Index

Jones, Heather

Original(1)

- Jones, Heather

Normalized(2)

- JONES, HEATHER

Auto-generated Synonyms(3)

- JONES, HEATHER

Indexed Value(4)

- **JONES, HEATHER**

Query

Brown, Heather

Original(5)

- Brown, Heather

Curated Synonyms

Brown, Heather
Jones, Heather

Normalized(6)

- BROWN, HEATHER

Name Variations(7)

- BROWN, HEATHER\b.*
- BROWN, HEATHER
- BROWN, H\b.*
- BROWN, H
- BROWN,

Auto-generated Synonyms (11)

- BROWN, HEATHER\b.*
- BROWN, HEATHER
- BROWN, H\b.*
- BROWN, H
- BROWN,

Expand w/ Index Synonyms (12)

- **BROWN, HEATHER\b.***
- **BROWN, HEATHER**
- BROWN, H\b.*
- BROWN, H
- BROWN,

Expand w/ Curated Synonyms (13)

- BROWN, HEATHER\b.*
- BROWN, HEATHER
- BROWN, H\b.*
- BROWN, H
- BROWN,
- JONES, H
- **JONES, HEATHER\b.***
- JONES,
- **JONES, HEATHER**

Original(8)

- BROWN, HEATHER
- JONES, HEATHER

Auto-generated Synonyms(9)

- BROWN, HEATHER
- JONES, HEATHER

Processed(10)

- **BROWN, HEATHER** => JONES\b., H, JONES\b., HEATHER\b.*, JONES\b.,, JONES\b., HEATHER
- JONES, HEATHER => BROWN\b., H, BROWN\b., HEATHER\b.*, BROWN\b., HEATHER, BROWN\b.,

Solr/Lucene Components

<fieldType name="author">

- Author Normalize filter
- Author Transliteration filter
- Author Query Variations filter
- Synonym filter

• ...

</fieldType>

<http://github.com/romanchyla/montysolr>

THANKS!

Jäy Luekèr

Rømåñ Čhÿla

Albérto Aæcōmäzzî



ADS 2.0 = <http://adslabs.org/adsabs>