

n Characters in Search of an Author

Improving Author Name Indexing & Searching for the ADS 2.0



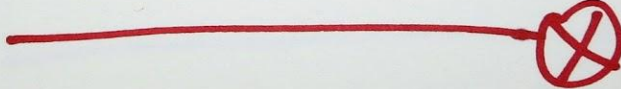


Waterwheels, Ace



Wäterwheels, Ace

Häcker, J



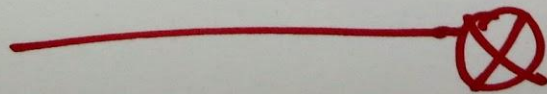
Hacker, J

Zoia Horn



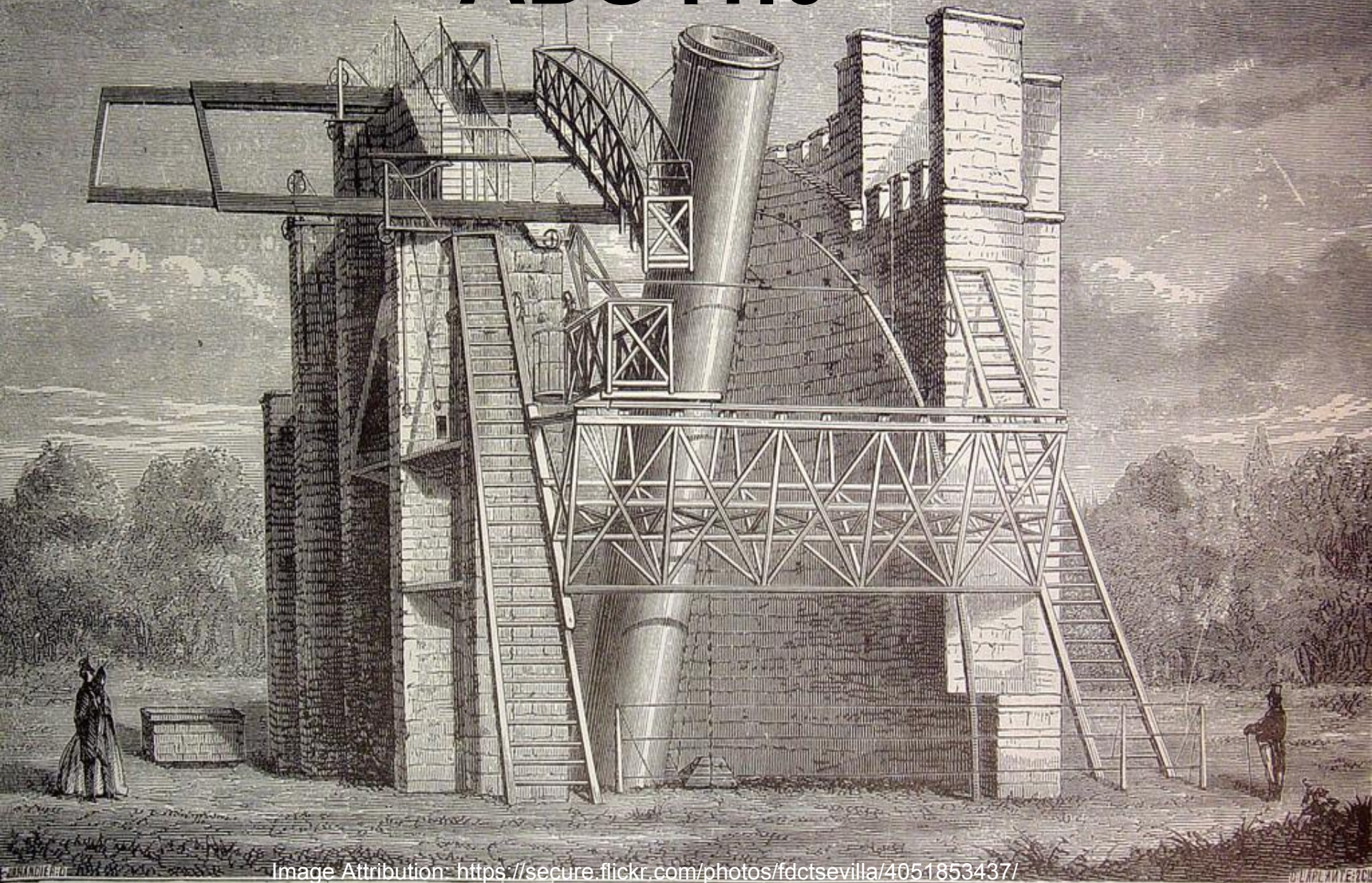
Horn, Z

Billz, Jimmy



Billz, James

ADS .1.0

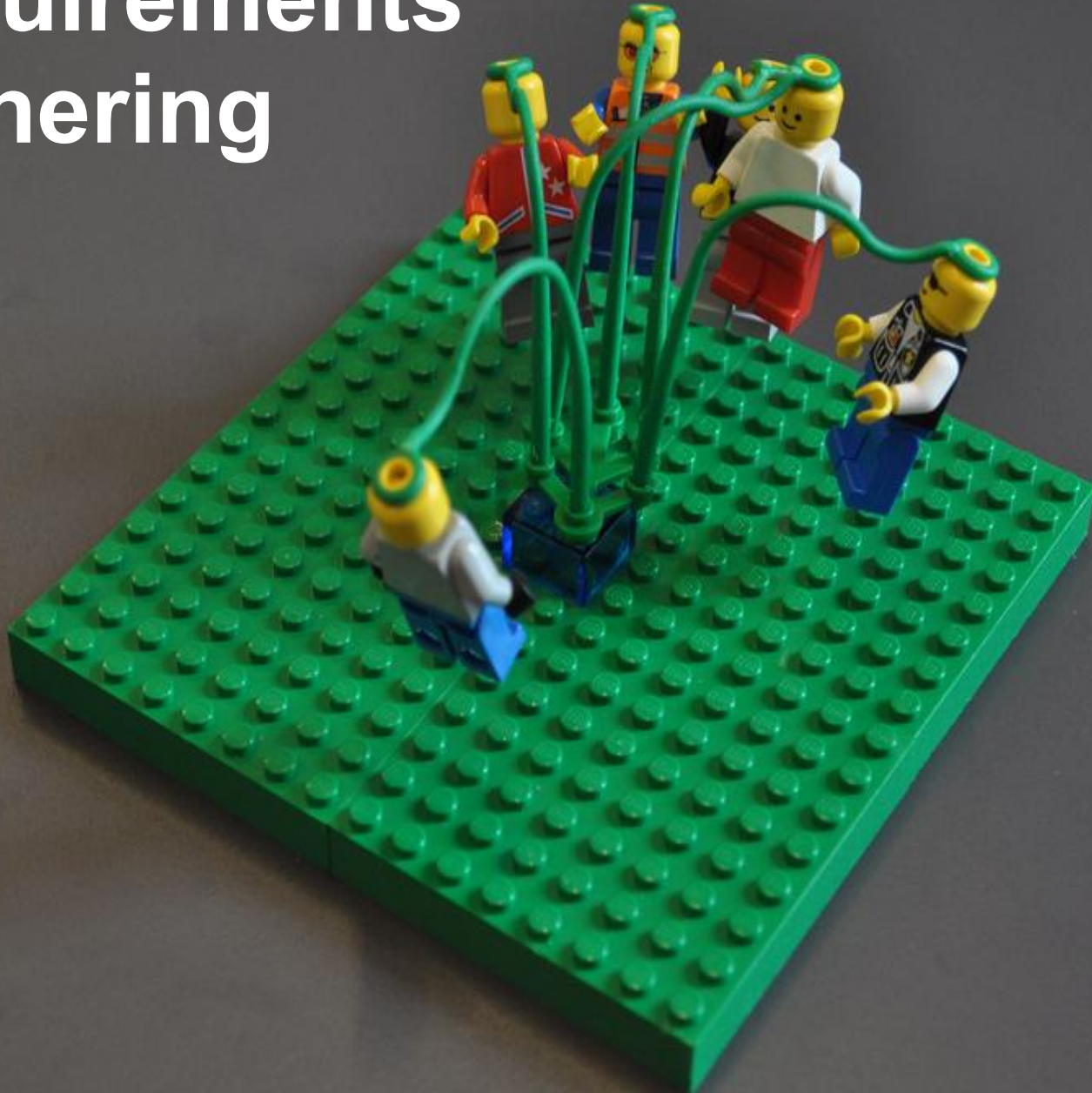


ADS 2.0



Image Attribution: http://www.nasa.gov/mission_pages/kepler/news/kepler20130103.html

Requirements Gathering



Req. #1



Hacker, J



Häcker, J

Req. #2



Hørn, Zoia

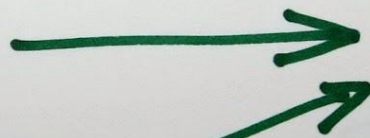


Horn, Zoia

Req. #3

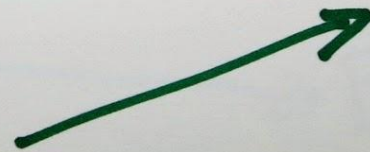


Waterwheels, Ace



Waterwheels, A

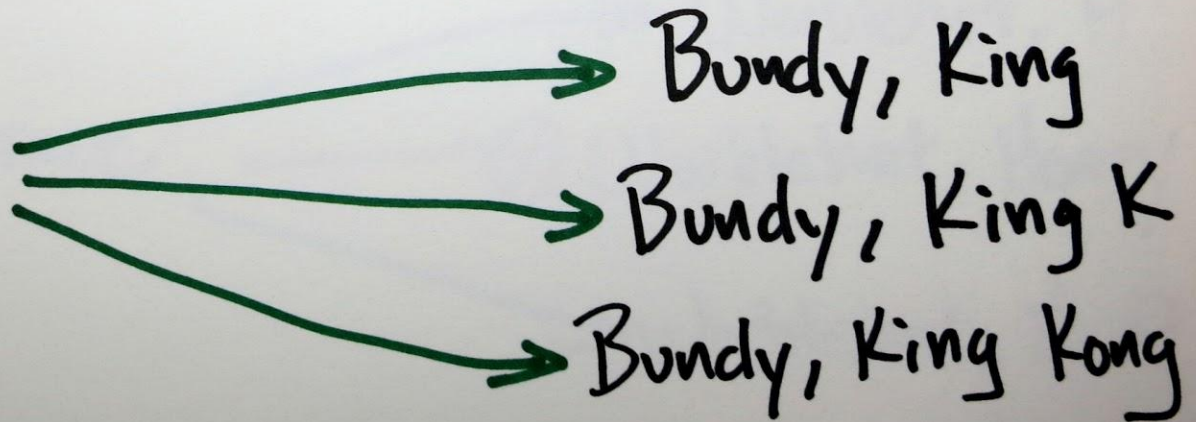
Waterwheels, Andy



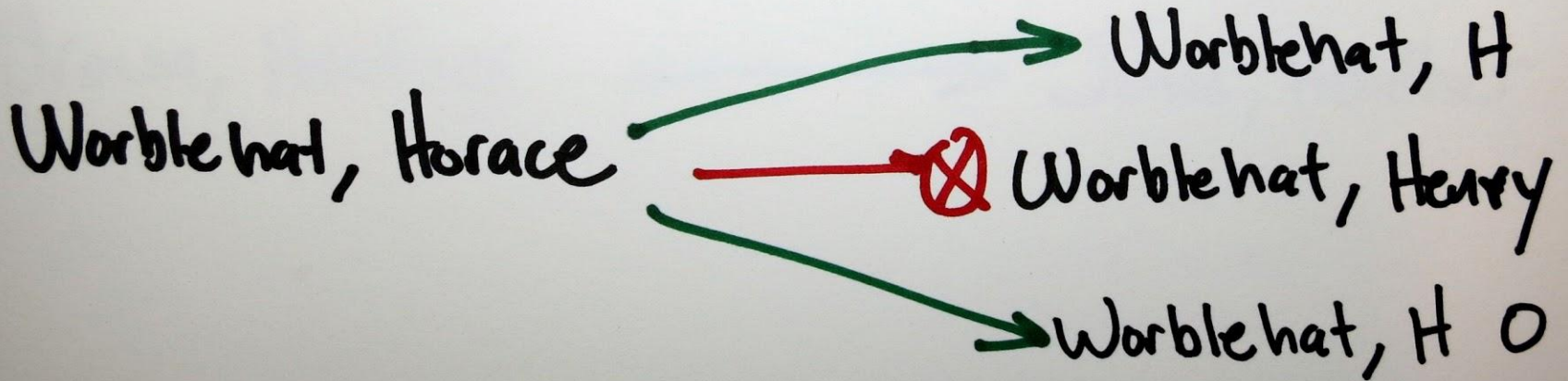
Req. #4



Bundy, K



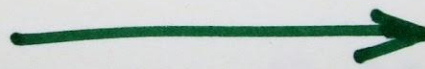
Req. #5



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, M M

Hørn, Zoia; → Hørn, Zoia

Brown-Jones, H → Brown Jones, H

Name Part Variations

Yogi, Maharishi M \rightarrow [Yogi, Maharishi M*
Yogi, M M*
Yogi, Maharishi
Yogi, M
Yogi,]

Transliteration (i)

Hübbard, M \longrightarrow [Hubbard, M
Huebbard, M]

č \longrightarrow [c, ch]

ø \longrightarrow [o, oe]

...

Transliteration (2)

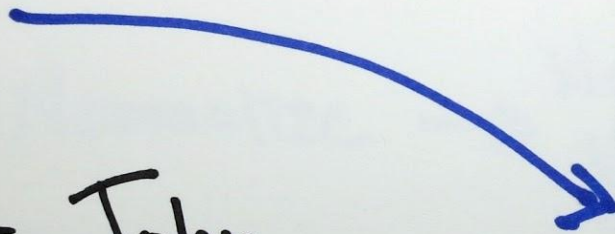
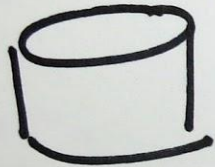
Hubbard, M

???

[Hübhard, M
Huebbard, M]

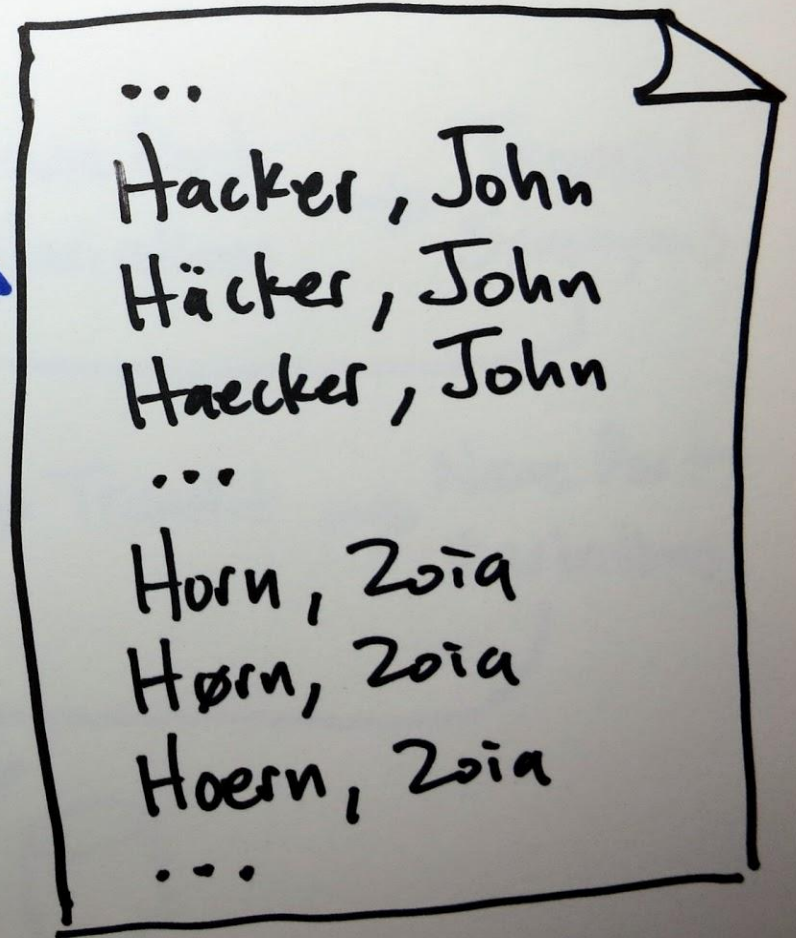
~~@takify~~

Index Introspection



Häcker, John

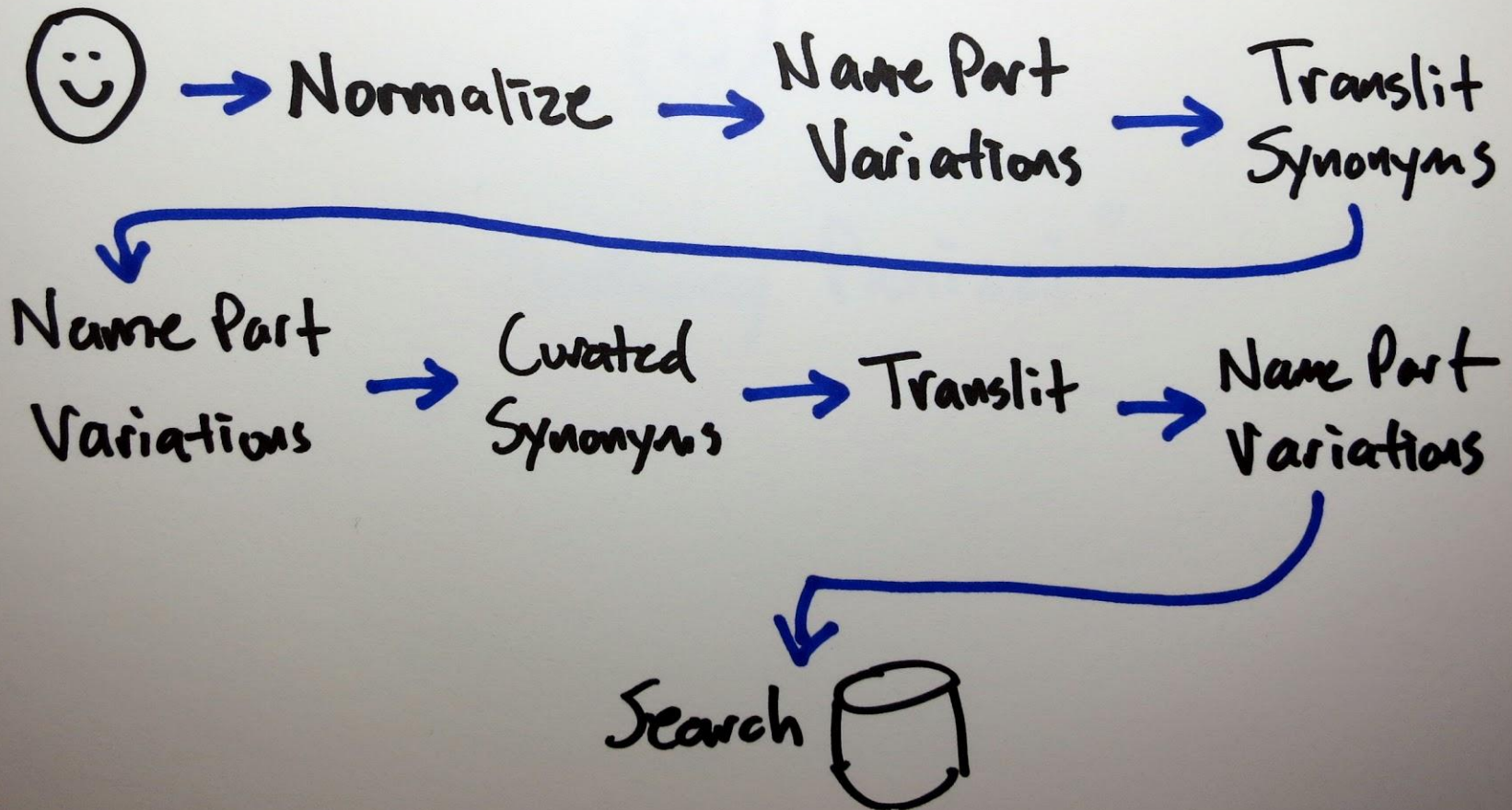
Hørn, Zoia



A vibrant, multi-colored nebula (likely the Helix or Ring Nebula) is centered in the frame against a dark, star-filled background. The nebula exhibits a mix of red, blue, and purple hues. The text "Query Expansion" is overlaid in white, bold, sans-serif font across the middle of the nebula.

Query Expansion

Walk through



User Input



"Anthony Panizzi"

Normalize



Parizzi, 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
 - ...
 - Krabs, Richard
 - Kraebs, Richard
 - Kräbs, Richard
 - ...

Name Part Variations



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

+ Paerizzi, Anthony *

+ Pänizzi, Anthony *

Panizzi, A

+ Paenizzi, A

+ Pänizzi, A

Panizzi, A *

+ Paenizzi, A *

+ Pänizzi, A *

Panizzi,

+ Paenizzi,

+ Pänizzi,

Curated Synonyms



Panizzi, Anthony

Paenizzi, Anthony

Pänizzi, Anthony

...

+ Pänizzi, Tony

...

Kräbs, Richard

Kräbs, Dickie L

...

Pänizzi, Anthony

Pänizzi, Tony

...

Transliteration



Panizzi, Anthony

Paenizzi, Anthony

Pänizzi, Anthony

...

+ Panizzi, Tony

+ Paenizzi, 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

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**

Curated Synonyms

Brown, Heather
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,

Solr/Lucene Components

<fieldType name="author">

- AuthorNormalizeFilter
- AuthorTransliterationFilter
- AuthorQueryVariationsFilter
- SynonymFilter

...

</fieldType>

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

THANKS!

Jäy Luekër

Rømän Čhýla

Albértø Aȝcōmäzzi



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