

WHAM! Use Cases

[Original Google docs link](#)

In this document, we will list the main use cases for auto-suggest and related examples.

Questions to consider:

- What kind of information should be used when generating matches for a query?
- Should certain types of matches be considered more important or ranked higher than others?
- What concrete examples can help define the answers to the above two questions as well as help us validate that the search results are returned as we expect?
- Do we include the "first last" version of a name as a variant for a primary label? If we do, what are the ramifications in terms of performance? There are over 6 million author names in the index and some percentage (?) of those are also subjects. The answer might depend on what kind of matching we support.

Types of matches

This list lays out some possibilities for matches but should not be considered a list of requirements.

- Full text match
- Partial text match
 - Match starts with same letters
 - Match contains same letters but in any position within the match
- If multi-word query, do all words show up in the same exact order somewhere within the match?
- If multi-word query, do all words show up in any order within the match?
- If multi-word query, use a combination of whole and partial word matches? (Match whole word on the first word; match partial on all ensuing words)

Use cases

Query	Relevant data in index	Matches displayed	Ranking comments (if applicable)
alb	label: "Einstein, Albert, 1879-1955" label: "Kleiner, Alberto Victoria and Albert Museum" label: "Alberti, Michael, 1682-1757" label: "Arkansas > Albert Pike Recreation Area" label: "Camus, Albert, 1913-1960 > Criticism and interpretation"	Einstein, Albert, 1879-1955 Kleiner, Alberto Victoria and Albert Museum Alberti, Michael, 1682-1757 Arkansas > Albert Pike Recreation Area Camus, Albert, 1913-1960 > Criticism and interpretation	
emil	label: "Dickinson, Emily 1830-1886" label: "Friedberg, Emil Albert, 1837-1910" label: "Emily binti Kaudon"	Dickinson, Emily 1830-1886 Friedberg, Emil Albert, 1837-1910 Emily binti Kaudon	
pear	label: "Buck, Pearl S. (Pearl Sydenstricker)" label: "Pearl, Raymond, 1879-1940" label: "Pearson, A. M. (Albert Marchant), 1916-" label: "China > Pearl River Delta" label: "Pearl Harbor, Attack on (Hawaii : 1941)"	Buck, Pearl S. (Pearl Sydenstricker) Pearl, Raymond, 1879-1940 Pearson, A. M. (Albert Marchant), 1916- China > Pearl River Delta Pearl Harbor, Attack on (Hawaii : 1941)	
emily di	label: "Dickinson, Emily 1830-1886" label: "Dicken, Emily F." label: "Dial-Driver, Emily"	Dickinson, Emily 1830-1886 Dicken, Emily F. Dial-Driver, Emily	
dickinson em	label: "Dickinson, Emily 1830-1886" label: "Dickinson, Emma" label: "Dickinson, Emmett"	Dickinson, Emily 1830-1886 Dickinson, Emma Dickinson, Emmett	

celtic grammar	label: "Celtic languages > Grammar, Comparative" label: "Celtic languages > Grammar, Historical" label: "Celtic languages > Grammar"	Celtic languages > Grammar, Comparative Celtic languages > Grammar, Historical Celtic languages > Grammar	
einstein albert	label: "Einstein, Albert, 1879-1955" label: "Einstein, Fred Albert" label: "Einstein, Albert Fred"	Einstein, Albert, 1879-1955 Einstein, Albert Fred Einstein, Fred Albert	
albert einstein	label: "Einstein, Albert, 1879-1955" label: "Einstein, Fred Albert" label: "Einstein, Albert Fred"	Einstein, Albert, 1879-1955 Einstein, Albert Fred Einstein, Fred Albert	
albert alister einstein	label: "Einstein, Albert, 1879-1955"	No matches	
einstein political views	label: "Einstein, Albert, 1879-1955 > Political and social views"	Einstein, Albert, 1879-1955 > Political and social views	
child care standards	label: "Child care services > Standards"	Child care services > Standards	
Query using variant	Relevant data in index	Matches displayed	Ranking comments (if applicable)
dzheyn edems	label: "Addams, Jane 1860-1935" variant_labels: ["Edems, Dzheyn, 1860-1935", "Addams, Laura Jane, 1860-1935"]	Addams, Jane 1860-1935	
c j smyth	label: "Smyth, Chris" variant_labels: [Smyth, C. J. (Chris J.)]	Smyth, Chris	
,	label: "Dickinson, Emily 1830-1886" variant_labels: ["Dikinson, mili, 1830-1886", "Dikinson, Emily, 1830-1886", "Ti-chin-sen, Ai-mi-li, 1830-1886", , , 18301886]	Dickinson, Emily 1830-1886	
	label: "Bronte, Emily 1818-1848" Variant_labels: ["Po-lang-te, Ai-mi-li, 1818-1848", ", 1818-1848", "Brontë, E. J. (Emily Jane), 1818-1848"]	Bronte, Emily 1818-1848	

Notes:

1. "albert einstein" and "einstein albert" (or "einstein, albert") should return the same results
2. matches are all at the beginning of words; no embedded substrings
3. multiple terms do not have to appear in order

Questions:

1. How to treat "and" and "or": do we throw them out of the query? For example, "einstein and religion" returns no suggestions but "einstein religion" does.

We could look into the usage of stop words in Solr, where words like "and" and "or" may effectively be ignored. In that case, both the examples would result in a match.

1. For variant or pseudonymous searches, do we include the search term in the response (for example, "Twain, Mark (Samuel Clemens)")?

Following the Wikidata lookup model, that seems reasonable for variant labels, where the preferred label would be listed first followed by the variant that matches what the user wrote. For pseudonyms, as we've discussed before, we'll need to use a different approach. In that case, the authority matching what the user wrote would be displayed first and a "see also" would indicate related pseudonyms.

Additional variant issues

We would want to prefer primary labels over variants.

What should the user experience be:

- when the variant is matched but the primary label looks different?
- When there are both primary label and variant matches for different entities?

Is it possible to require matches after the first word with partial matches following?

Should we require “stricter” matches (i.e. whole word matches with variants)?

Pseudonym scenarios

Questions:

Wikidata has a pseudonym property which returns literals (for pseudonyms) for the person. This information could be retrieved for search purposes to enable matching on pseudonyms as well. LCNAF uses “see also” properties that may or may not be

User query	Data	Behavior (Generally, result selection should lead to search in appropriate field which is more flexible than facet search)
Samuel Clemens	Samuel Clemens: -separate authority in catalog -has distinct URI Mark Twain: -separate authority -has distinct URI	Show “Samuel Clemens (#)”, with connection to pseudonym “See also Mark Twain(%)” and allow selection of that item as well
Street liberty (made up example)	Liberty Mutual: <ul style="list-style-type: none">• Authority in catalog• Distinct URI Street liberty: <ul style="list-style-type: none">• Authority NOT in catalog but authority exists in LCNAF• Distinct URI	Show “Liberty Mutual (#) (Street liberty)” (?) indicating pseudonym match No need to show separate connection to “Street liberty” pseudonym b/c it does not exist as a separate authority (i.e. separate search with pseudonym not required)
Fidelity stocks	Temperamental oddities <ul style="list-style-type: none">• Authority in catalog• Distinct URI Fidelity stocks: <ul style="list-style-type: none">• Authority does not exist at all• No URI	Show “Temperamental oddities (#) (Fidelity stocks)” indicating pseudonym match No separate matches/URIs to take into account
Fictional Physicist	A joint pseudonym for multiple people	Show ? Joint pseudonym first? There is more than one primary label in this case

Q: Is there such a thing as a primary identity in lcnaf to begin with?

If “see also” can go in any direction, then either one could be considered the “primary” label?

So if the user types in samuel clemens, they see that plus a “see also” pointing to mark twain

If they type in “mark twain”, they see that as the primary authority, with “see also” pointing to samuel clemens