# InquiryMap: Automatic Labeling of Libchat Initial Questions

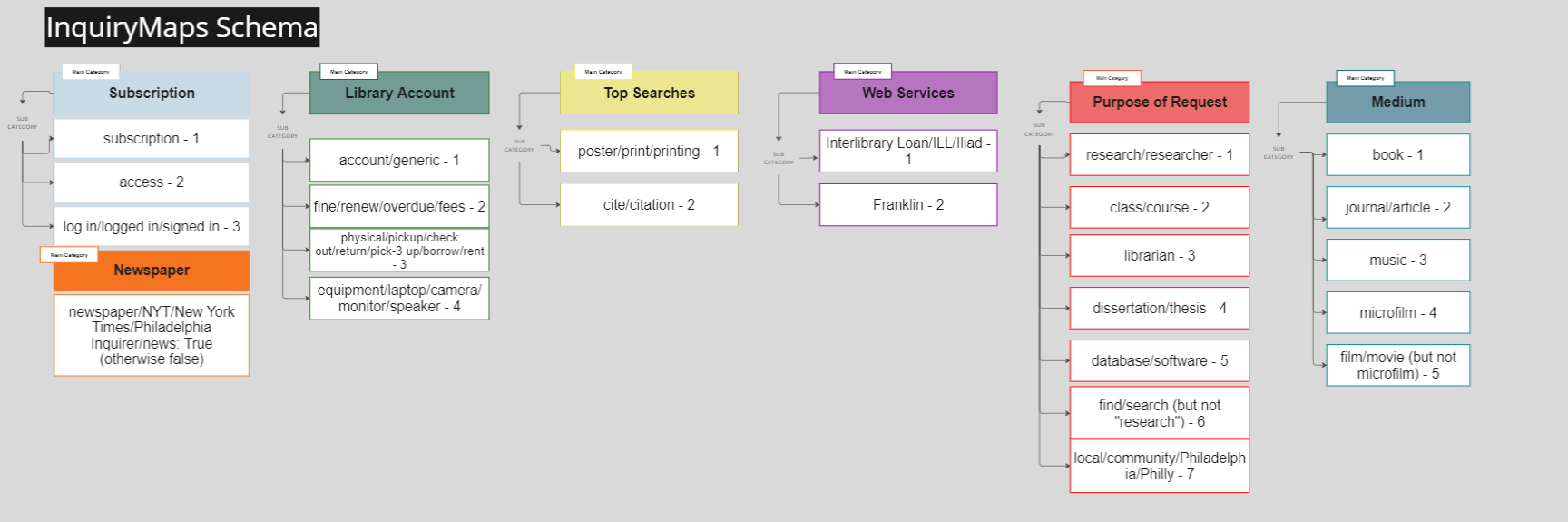
InquiryMap is a service designed to enhance the library chat data with automatically labelled content, sentiment, and affiliation tags. We describe each of the tags in the following document.

Content flags:

The content tags arise from a schema of 7 main categories and 24 subcategories derived by UX Researcher Virshae Campbell and Bollinger Fellow in Innovation Angelica Rivera. The first 4 categories are from a “by eye” examination of a set of chats, as well as from the most frequently searched terms on the library website. The last 3 categories were created to capture as many of the remaining chats as possible. These categories and numbers were derived from the 1,924 Libchats from 5/01/2023-05/31/24.

**Note that a key feature of this schema is that multiple categories can be assigned at once!**

Out of 1,924 chats, 77% get labelled from the keywords, an additional 13% can be easily be identified by eye with the schema below, and the remaining 10% are miscellaneous. The missing 13% occurred due to user typos or not searching for all possible cases of a word in the original test. **We highlight that a total of 90% of chats can be labeled with our schema.** The subcategories are assigned by searching for particular words or phrases within the initial chat questions, these words/phrases are given in the diagram below.




The value of this type of categorization is demonstrated below, using the following flags:

| Medium | Purpose of Search | Services | Account _Q | Top Searches | Subscriptions | Newspapers |
|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 1 | 2 | 2 | False |

we can infer that the user's chat is **about research or from a researcher about an account for a subscription service that involves citations.** Below is the actual user statement:

"I would like to gain access to the premium features on Web of Science (e.g. search all citations of a paper). As a CHOP researcher, working under ***, I have access to Penn Libraries. Do I also have access to the Web of Science features?"

**Without looking at the actual chat we were able to determine precisely what the user needed help with, protecting privacy and extracting insights!**

Affiliation Flags:

The "user type" column indicating whether the question was asked by an Alum, a Visitor a student, or a faculty/staff member. Am currently identifying by the use of the following indicators:

1) The presence of those identifier strings ("visitor", "student", "researcher", etc.) in the initial question.
2) Using the contact information provided or the pennkey to see if there is a match to our Alma demographics. This can provide the identifiers above.

Additionally, we can populate a "school" column where we have matches to the Alma demographics table. The demographics will not be 100% populated (users are not required to self-identify by pennid or email), but the information gathered could be useful for determining which populations experience pain points.

Sentiment flag:

We used the VADER sentiment analysis tool to gain insight into user sentiments. VADER is trained on social media posts, with a (human-assigned) lexicon designed to handle colloquial input such as emojis and internet/chat shorthand (such as "lol"). It is also capable of recognizing mixed sentiment in a single

sentence. We have used VADER solely on the user lines of the chat, averaging the values for each line to get a final “overall sentiment” score. The breakdown of sentiment scores into sentiment bins is as is follows:

Negative: Compound < -0.05
Neutral: -0.05 < Compound < 0.05
Positive: Compound > 0.05

**A note of Caution**: the user may simply fail to respond to the librarian in the chat (i.e.-they get the answer they were looking for and just leave), or may get disconnected, so neutral scores should not be read as necessarily being unsuccessful or negative.

We have reviewed a subset of the labeled chats by hand to ensure chats were labeled with the sentiments we expected. We found an 83% success rate in the automatic labeling of sentiments. We found that an average compound score of >=0.3 almost always corresponds with the question being solved, next steps identified, and the user being happy (false positive rate of only 5% in the 2023-2024 academic year data, whereas for the “positive” category the false positive rate is 9%), so we form an additional subset with this score threshold, considered “ecstatic”. We discovered a coding artifact from utilizing VADER: “positive” chats are often longer with multiple questions being asked or troubleshooting being required. The types of question that this tag normally describes include queries such as “Can you help me run this software?” On the other hand, the majority of chats in the “ecstatic” category are short “easy questions”, such as “When is the library open?”