# YouUndestood.me

YUM is an online environment built around a search engine, intended to make the search process valuable for children. Yum is not meant to be treated as a new, child-oriented search engine, since studies \cite{ref} show that children tend to prefer popular search engines such as Google to perform the information-seeking tasks. Instead, YUM acts as an intermediate layer between the child and an existing search engine in order to facilitate the communication between the two of them. For doing so, YUM puts into practice strategies oriented to address issues children face when using popular search engines, as well as strategies that aim to enhance the search experience in order to foster learning. A description of the mentioned strategies is provided below.

## Search intent

Children are unusually successful in formulating succinct queries \cite{Bil11}. Studies show that they tend to write natural language queries, instead of short, keyword-based ones that search engines usually expect \cite{Dru09}. Unfortunately, the longer the query, the less likely it is for a search engine to retrieve relevant results in response to it, a fact that can lead children to frustration because of the inability to complete their information seeking tasks\cite{Dru09}. In addition, children also tend to misspell words, misspellings that can differ from the ones an average adult does. For example, children commonly repeat letters in a word to emphasize it, such as in “amaaaaaaaaazing”. In order to best satisfy children needs Yum takes advantage of QuIK\cite{Sven} a search intent module specifically designed for children. QuIK is able to address common patterns in queries written by children including but not limited to: diminutives, emphasis, children trendy terms or children specific misspellings. In doing so, QuIK transforms an initial child-query into a new keyword based query that captures the information expressed by the child in a way that can more easily be comprehended by search engines.

## Query suggestion

Even if the search intent module can fix and extract the main intent of a number of queries, there are some, such as ambiguous queries or queries that have different meaning depending on the user’s interest that it cannot. In both cases, the user, as the main stakeholder, is the only one that knows the purpose of his search and can therefore improve the query in order to better inform the search process, based on the goal of his respective information seeking task. For this reason, Yum takes advantage of ReQuIK \cite{requik}, a query recommender that is specifically tailored to children. This tool provides suggestions that extending or shortening the query in diverse ways and potentially containing different meanings depending on the point of view of the user. Those suggestions are given to the user, so that he can pick the one that best fits his information requirements. ReQuIK is a multi-criteria recommendation system based on traits commonly associated with children that suggests queries that are associated with children topics, will lead to the retrieval of resources with levels of readability matching those of the K-12 audience, and are diverse enough to capture the different topics children can be interested in.

## Filtering by readability

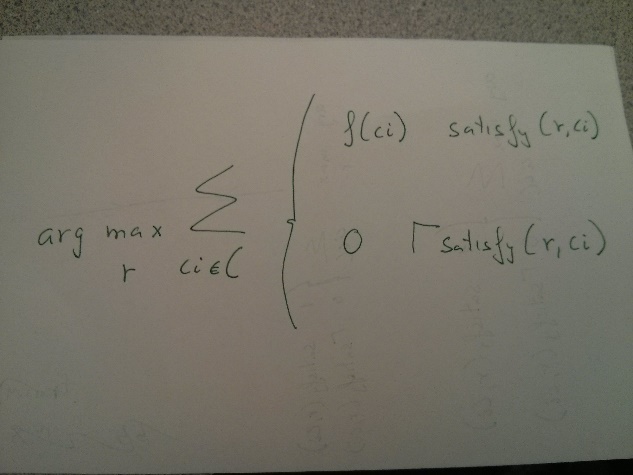
Even when the search engine has understood the intent of a child query and retrieves results that match the information needs of the corresponding users, suitability of retrieved resources is still not assured. K-12 students tend to find difficult to understand documents containing complex or technical vocabulary. For example, in the case where a child is looking for information about chemistry, retrieving a scientific publication would not be adequate, while retrieving information from an elementary chemistry book would. On cases where all retrieved documents are too complex, children may not succeed in completing their information seeking process. In order to avoid this situation, Yum incorporates as part of its environment a filtering strategy based on readability levels which ensures that the retrieved documents will match the reading ability of each individual user. Yum allow users to go through a one-time process where they can select their grade level. This grade level determines their readability level, which is matched against the one of documents retrieved by the search engine, eliminating the ones that are not in one grade level range from the one of the student. For predicting the readability of retrieved resources Yum uses the Flesh-Kincaid \ref{Fle48} (see formula XX) readability formula. We initially selected this formula given that it is widely used for measuring readability of texts among educators nationwide and is considered as a standard by institutions\ref{porposal31\_38\_40} to measure readability levels. However, we are aware that it is only based on simple text-based features, which is why we plan to improve this formula in the future (see Future Work Section).

Placeholder for fleshkincaid

## Tracking

K-12 students have diverse reading abilities. Even in same grade class, students’ readability skills can differ. Furthermore, the reading skills of each individual progressively improve over time\cite{sh13}. Consequently, a one-size-fits-all strategy, is not applicable for conducting successful information-seeking tasks that lead to the retrieval of resources individual users can read and understand. For this reason, Yum allows its users to provide optional feedback determining whether the resource was too “*easy”*, “*OK”* or “too *complex”* for them. This feedback is used to determine and update the reading skills for each user so that the system can retrieve documents adequate to that level of readability.

The problem of students’ readability prediction and updating can be seen as a constraint satisfaction maximization problem, where each feedback **f** produces a new constraint that needs to be satisfied by the readability of the student **s.** For example, a student giving a feedback of “too hard” to a document of readability level 5 would generate the constraint readability(**s**) < 5 stating that the readability of student **s** should be lower than 5. The predicted readability is the one that maximizes the amount of constraints satisfied. In case of multiple maximum values, the one that has the biggest distance to its two (lower and upper) closest constraints is selected. Note that we also consider the time the resource was read by the student, decreasing the importance of the constraint the older it was. This permits Yum to prioritize recent feedback, and allow the readability prediction adapt to the student over time.



Where **r** is the readability value, **ci** is an individual constraint among the set C of constraints created based on the feedback provided on retrieved resources by a user **s**, and **f(ci)** refers to how new the constraint is, starting at 9 is the constraint was created at this month, and getting reduced by one if for every month, ending in 0 if the constraint was created outside the academic year.

Furthermore, in order to avoid the cold start problem, where the student does not have any constraint to satisfy, every student account is initialized with two constraints based on the grade level of the student. Those constraints are representative of one grade of deviation from the current grade of the student: readability(**s**) < grade(**s**)+1 and readability(**s**) > grade(**s**) -1. These constraints give a starting point to Yum, that will eventually be better adjusted when the student start using the environment. This initialization takes place at the beginning of each academic year, however previous academic year readability prediction is used instead of the grade level if this data is available.

# Yum for teachers

Children are not the only stakeholders of Yum, since teachers can also benefit of using it within the class environment. Work setting standards have changed from a vertical structure, where only the top individuals of the pyramid had to think critically and the lower parts just followed directions, to an horizontal structure, where each individual is expected to collaborate with others and solve important problems using identification, searching, synthetizing, and communication skills \ref{leu13}. Given this change, education plans oriented to meet the new requirements of the current industry, such as the Common Core State Standards (CCSS) Initiative, have been developed. CCSS requests educators to make an emphasis on higher level thinking during reading and writing and focus on the acquisition of skills such as research and comprehension using digital tools, such as search engines\ref{leu13}. Furthermore, educational studies \ref{kni15} showcase the benefit of in class exercises such as exploratory talks, where students are asked to resolve a problem in groups discussing information found in resources obtained using a search engine. Unfortunately, teachers might not be able to propose such a task to their students and lead critical discussions, if students have problems using search engines, whether that be struggling to find the right query or not being able to understand the retrieved documents due to their complexity. YUm helps the teacher overcome those issues in class and focus on the discussion, rather than on the manner in which students should formulate queries? Or the type of results they access. Furthermore, YUm serves the teacher as a monitoring tool that allows to check how students are doing based on the resources they have retrieved and given feedback so that he can give them assistance when needed. Therefore, we think that YUm will not only facilitate learning when children use it for their information discovery assignments at home, but it will also help teachers the classroom environment.

@article{sh13,

title={Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling},

author={Shin, Tacksoo and Davison, Mark L and Long, Jeffrey D and Chan, Chi-Keung and Heistad, David},

journal={Learning and Individual Differences},

volume={23},

pages={92--100},

year={2013},

publisher={Elsevier}

}

@article{leu13,

title={The new literacies of online research and comprehension: Assessing and preparing students for the 21st century with Common Core State Standards},

author={Leu, Donald J and Forzani, Elena and Burlingame, Cheryl and Kulikowich, Jonna and Sedransk, Nell and Coiro, Julie and Kennedy, Clint},

journal={Quality reading instruction in the age of common core standards},

pages={219--236},

year={2013}

}

@article{kni15,

title={The role of exploratory talk in classroom search engine tasks},

author={Knight, Simon and Mercer, Neil},

journal={Technology, Pedagogy and Education},

volume={24},

number={3},

pages={303--319},

year={2015},

publisher={Taylor \& Francis}

}

@incollection{Bil11,

title={Evaluating leading Web search engines on children's queries},

author={Bilal, Dania and Ellis, Rebekah},

booktitle={Human-Computer Interaction. Users and Applications},

pages={549--558},

year={2011},

publisher={Springer}

}

@inproceedings{Dru09,

title={How children search the internet with keyword interfaces},

author={Druin, Allison and Foss, Elizabeth and Hatley, Leshell and Golub, Evan and Guha, Mona Leigh and Fails, Jerry and Hutchinson, Hilary},

booktitle={SIGCHI},

pages={89--96},

year={2009},

organization={ACM}

}

@inproceedings{Sven,

title={``Is Sven Seven?": A Search Intent Module for Children},

author={Dragovic, Nevena and Madrazo, Ion and Pera, Maria Soledad},

booktitle={ACM SIGIR},

pages={},

year={2016}

}

@article{Fle48,

title={A new readability yardstick.},

author={Flesch, Rudolph},

journal={Journal of Applied Psychology},

volume={32},

number={3},

pages={221},

year={1948},

publisher={American Psychological Association}

}