Provides the interface between the person doing the searching and the search engine Architecture of a search engine Improves the initial query **G** query Carlos Andrés Reyes Evangelista | August 25, 2019 Constructs the final Query input Results output list of documents ısing scores based on a retrieval model Uses the User interaction structures fron Builds the indexing process structure that to return a ranked enable searching list of documents to the user's query Design of ranking algorithms and the asociated indexes to decrease response time and increase query throughput Indexing process Query process Ranking that potentially can be reuse Major functions Effectiveness Query process Users' queries and their interactions helps to improve Logging queries, advertising and find relevant documents Architecture of a Primary goals search engine Ranking analysis Log data and explicit relevance judgements are used to measure the Evaluation effectiveness of a ranking algorithm Speed Efficiency Performance To identify and make analysis available the documents that will be searched. Involves monitoring and improving overall Parts of a document that are stored in the index Often, it will require system performance Creates the indexes building a collection by Processes and used in searching. The or data structures crawling or scanning the the tokens set of all the terms is that enable fast source of information. Index terms Parser called index vocabulary searching from the output of the text Transforms Creates a document A B C D documents data store, which into: Text transformation documents its text and Feature metadata Part of a text document These indexes are → Play Stemming Index creation that is used to represent supposed to be efficiently updated its content, which also describes an index term when new documents are Groups Classifier Document Information Link words Crawler Conversion data store extraction derived from Stopping extraction a common and analysis - IXI Weighting stem *** distribution labels to Identifies documents or Follow the links Database that ill Ø □ □ → Ø Ø □ Removes terms that parts of Provides the on the web pages stores large common are more documents Document feeds Documents are search engine to discover and numbers of complex than with a rating of are a mechanism converted to plain download new a word documents and popularity text and to a for accessing a unstructured data pages. real-time stream consistent Gathers and Reflects the Changes from Distributes of documents. encoding records statistical relative document-term indexes across information about importance of information to multiple nodes for words, features term-document enable words in and documents information parallelism documents