Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Evaluation of Quantum Computing for IR and RS

Nicola Ferro

@frrncl

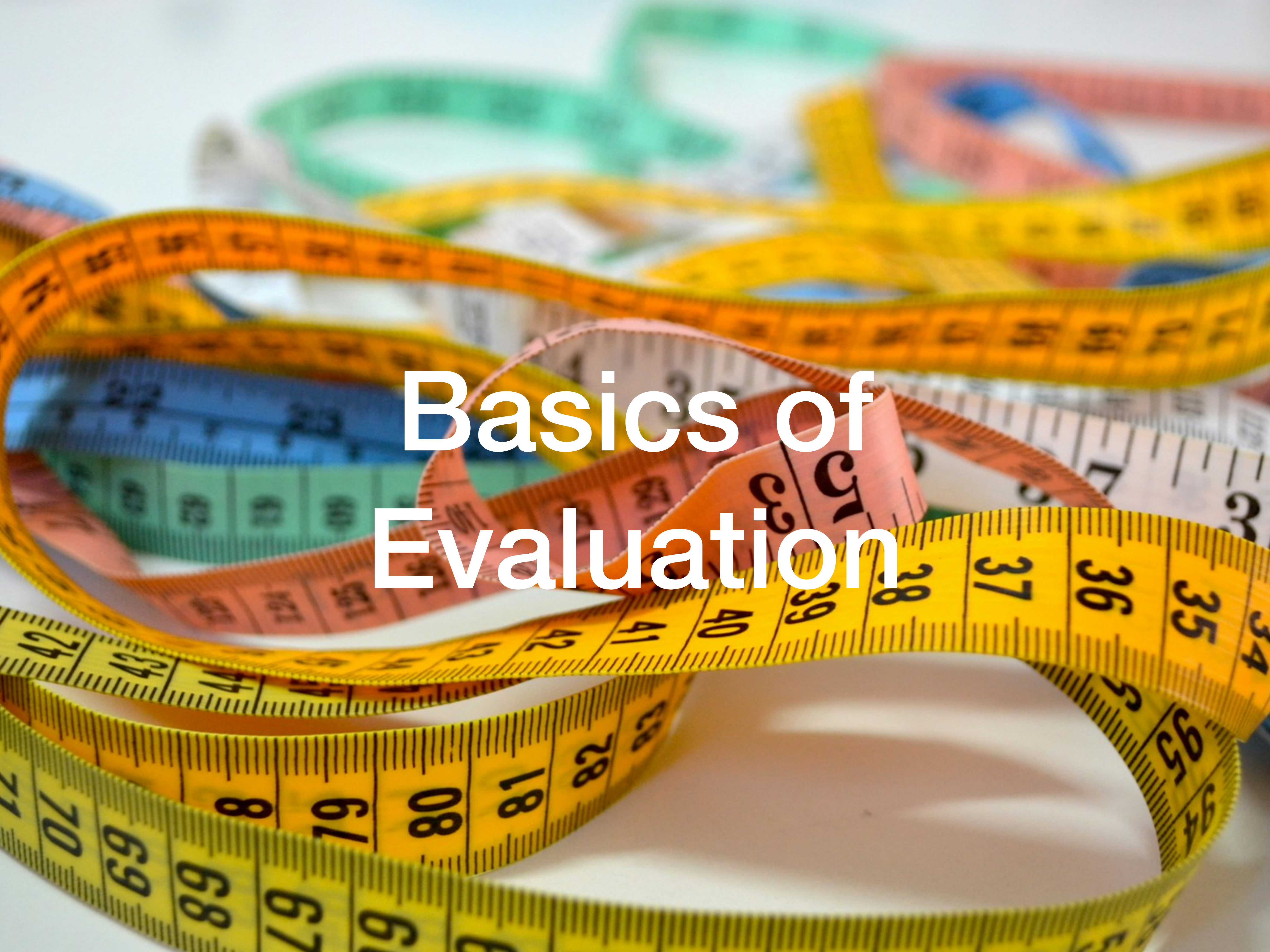
Intelligent Interactive Information Access (IIIA) Hub
Department of Information Engineering
University of Padua

Tutorial on Quantum Computing for Information Retrieval and Recommender Systems
ECIR 2024, 24th March 2024, Glasgow, UK

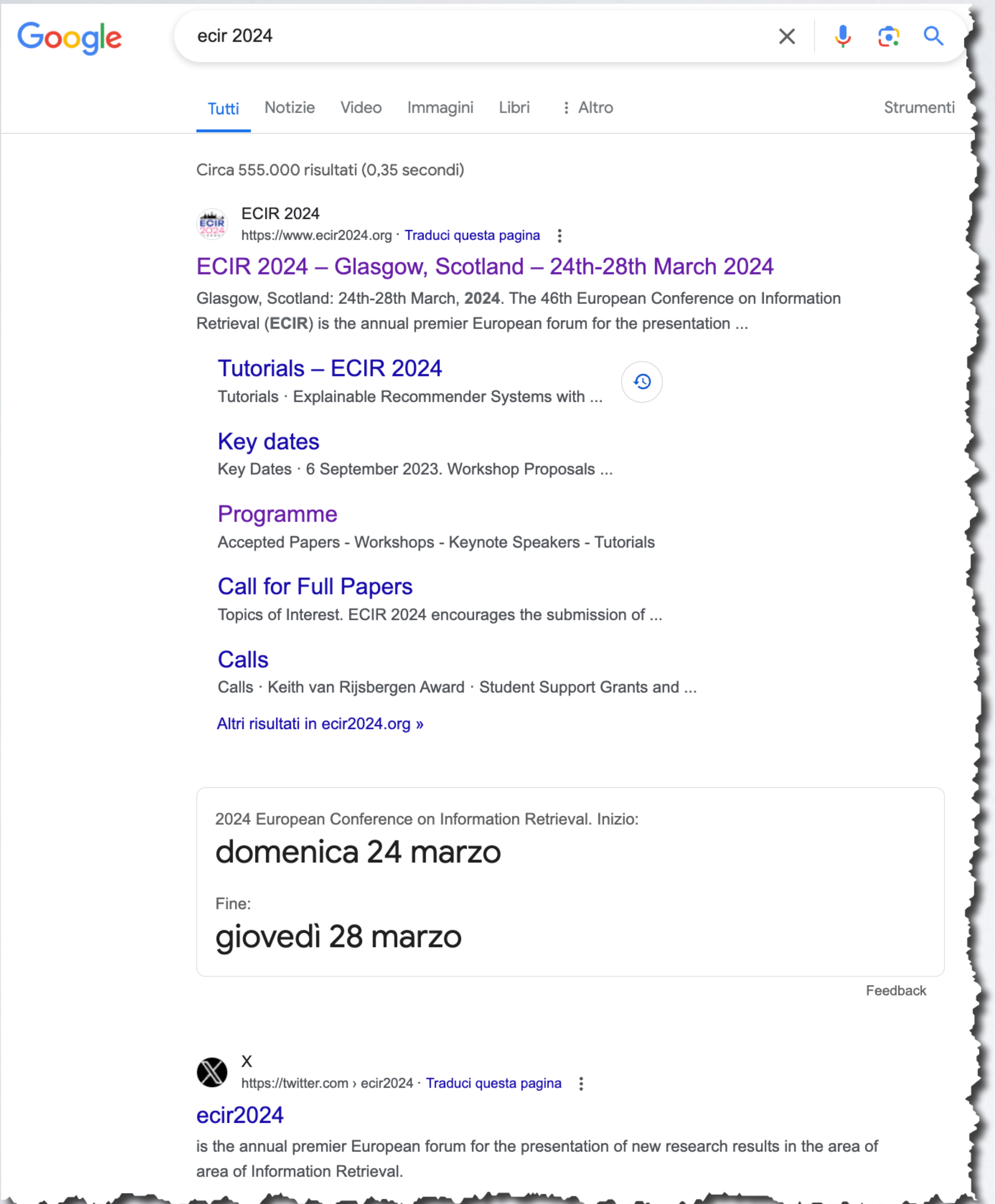
Outline

- Basics of Evaluation
- QuantumCLEF

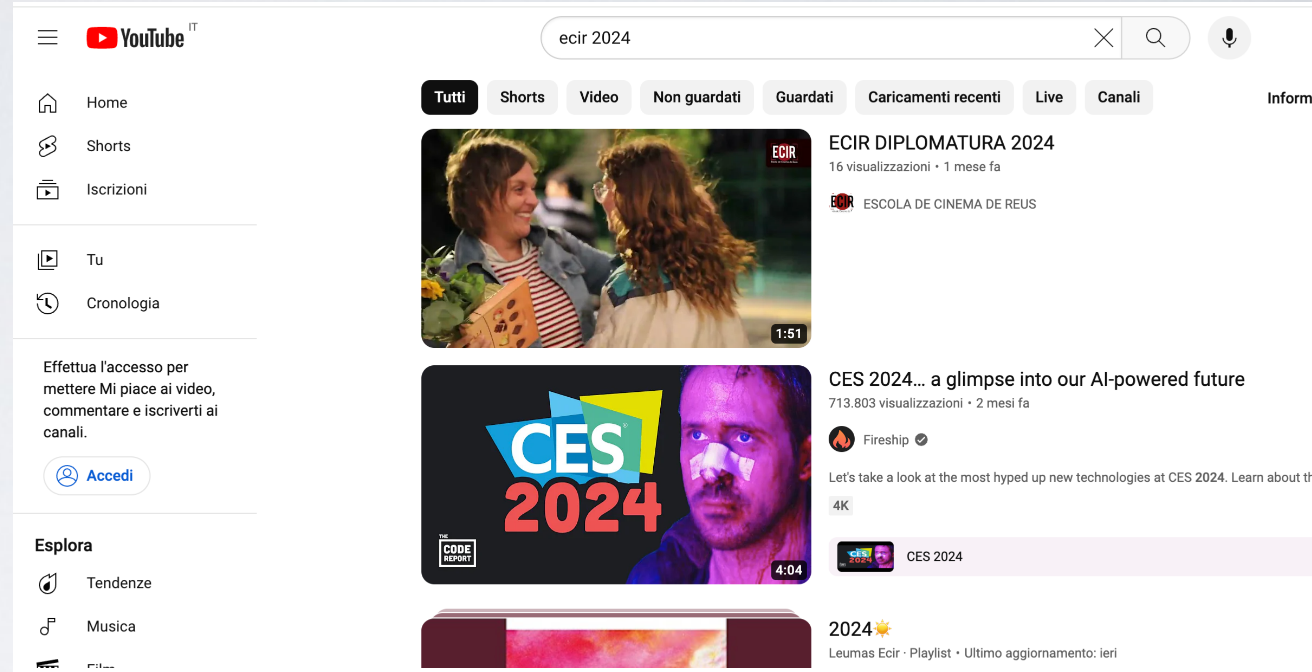
Basics of Evaluation



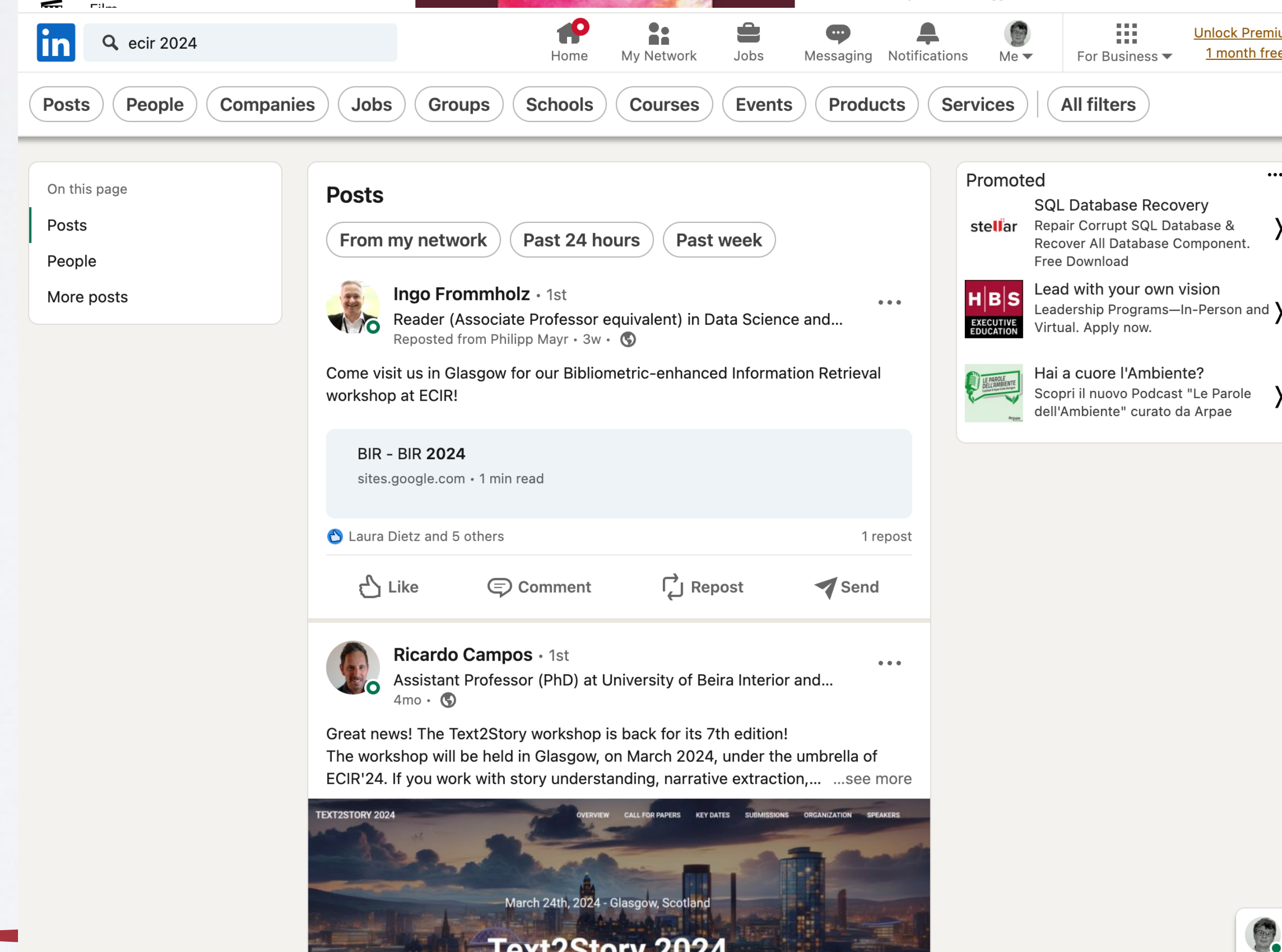
Our Goal



Google
ecir 2024
Tutti
Notizie
Video
Immagini
Libri
Altro
Strumenti
Circa 555.000 risultati (0,35 secondi)
ECIR 2024
https://www.ecir2024.org · Traduci questa pagina
ECIR 2024 – Glasgow, Scotland – 24th-28th March 2024
Glasgow, Scotland: 24th-28th March, 2024. The 46th European Conference on Information Retrieval (ECIR) is the annual premier European forum for the presentation ...
Tutorials – ECIR 2024
Tutorials · Explainable Recommender Systems with ...
Key dates
Key Dates · 6 September 2023. Workshop Proposals ...
Programme
Accepted Papers - Workshops - Keynote Speakers - Tutorials
Call for Full Papers
Topics of Interest. ECIR 2024 encourages the submission of ...
Calls
Calls · Keith van Rijsbergen Award · Student Support Grants and ...
Altri risultati in ecir2024.org »
2024 European Conference on Information Retrieval. Inizio:
domenica 24 marzo
Fine:
giovedì 28 marzo
Feedback
X
https://twitter.com › ecir2024 · Traduci questa pagina
ecir2024
is the annual premier European forum for the presentation of new research results in the area of Information Retrieval.

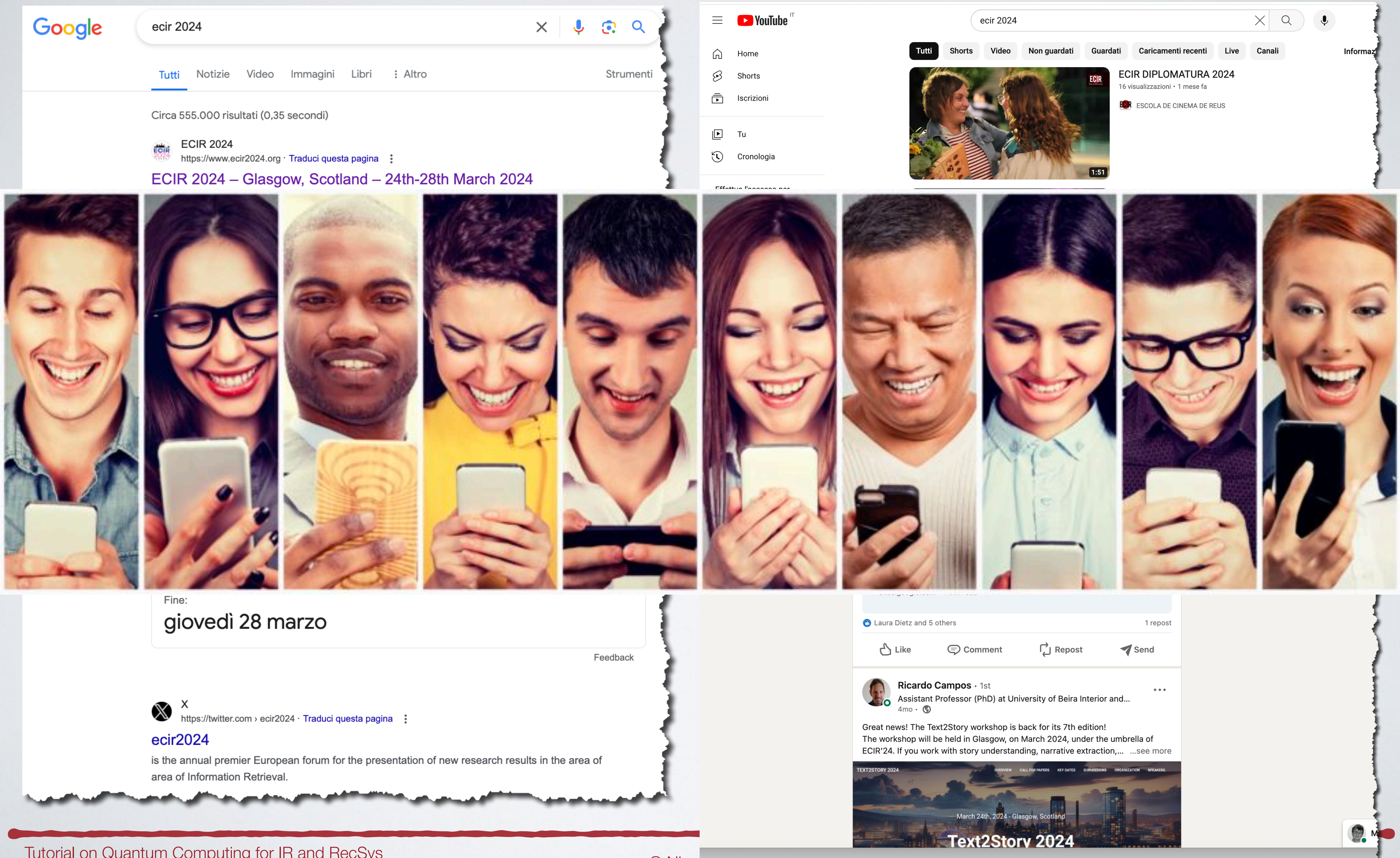


YouTube IT
ecir 2024
Home
Shorts
Iscrizioni
Tu
Cronologia
Effettua l'accesso per mettere Mi piace ai video, commentare e iscriverti ai canali.
Accedi
Esplora
Tendenze
Musica
Tutti
Shorts
Video
Non guardati
Guardati
Caricamenti recenti
Live
Canali
ECIR DIPLOMATURA 2024
16 visualizzazioni · 1 mese fa
ESCOLA DE CINEMA DE REUS
1:51
CES 2024… a glimpse into our AI-powered future
713.803 visualizzazioni · 2 mesi fa
Fireship
Let's take a look at the most hyped up new technologies at CES 2024. Learn about the
4K
CES 2024
4:04
2024
Leumas Ecir · Playlist · Ultimo aggiornamento: ieri



ecir 2024
Home
My Network
Jobs
Messaging
Notifications
Me
For Business
Unlock Premium
1 month free
Posts
People
Companies
Jobs
Groups
Schools
Courses
Events
Products
Services
All filters
On this page
Posts
People
More posts
Posts
From my network
Past 24 hours
Past week
Ingo Frommholz · 1st
Reader (Associate Professor equivalent) in Data Science and...
Reposted from Philipp Mayr · 3w ·
Come visit us in Glasgow for our Bibliometric-enhanced Information Retrieval workshop at ECIR!
BIR - BIR 2024
sites.google.com · 1 min read
Laura Dietz and 5 others
1 repost
Like
Comment
Repost
Send
Ricardo Campos · 1st
Assistant Professor (PhD) at University of Beira Interior and...
4mo ·
Great news! The Text2Story workshop is back for its 7th edition!
The workshop will be held in Glasgow, on March 2024, under the umbrella of ECIR'24. If you work with story understanding, narrative extraction,... ...see more
March 24th, 2024 - Glasgow, Scotland
Text2Story 2024
Promoted
SQL Database Recovery
Repair Corrupt SQL Database & Recover All Database Component. Free Download
Lead with your own vision
Leadership Programs—In-Person and Virtual. Apply now.
Hai a cuore l'Ambiente?
Scopri il nuovo Podcast "Le Parole dell'Ambiente" curato da Arpae

Our Goal



Google
ecir 2024
Tutti
Notizie
Video
Immagini
Libri
Altro
Strumenti
Circa 555.000 risultati (0,35 secondi)
ECIR 2024
https://www.ecir2024.org · Traduci questa pagina
ECIR 2024 – Glasgow, Scotland – 24th-28th March 2024
Fine:
giovedì 28 marzo
Feedback
X
https://twitter.com › ecir2024 · Traduci questa pagina
ecir2024
is the annual premier European forum for the presentation of new research results in the area of area of Information Retrieval.
YouTube
ecir 2024
Home
Shorts
Iscrizioni
Tu
Cronologia
Tutti
Shorts
Video
Non guardati
Guardati
Caricamenti recenti
Live
Canali
ECIR DIPLOMATURA 2024
16 visualizzazioni · 1 mese fa
ESCOLA DE CINEMA DE REUS
1:51
Laura Dietz and 5 others
1 repost
Like
Comment
Repost
Send
Ricardo Campos · 1st
Assistant Professor (PhD) at University of Beira Interior and...
4mo ·
Great news! The Text2Story workshop is back for its 7th edition!
The workshop will be held in Glasgow, on March 2024, under the umbrella of ECIR'24. If you work with story understanding, narrative extraction,... ...see more
TEXT2STORY 2024
March 24th, 2024 - Glasgow, Scotland
Text2Story 2024

Why Evaluation?

"To measure is to know"

"If you cannot measure it,
you cannot improve it"

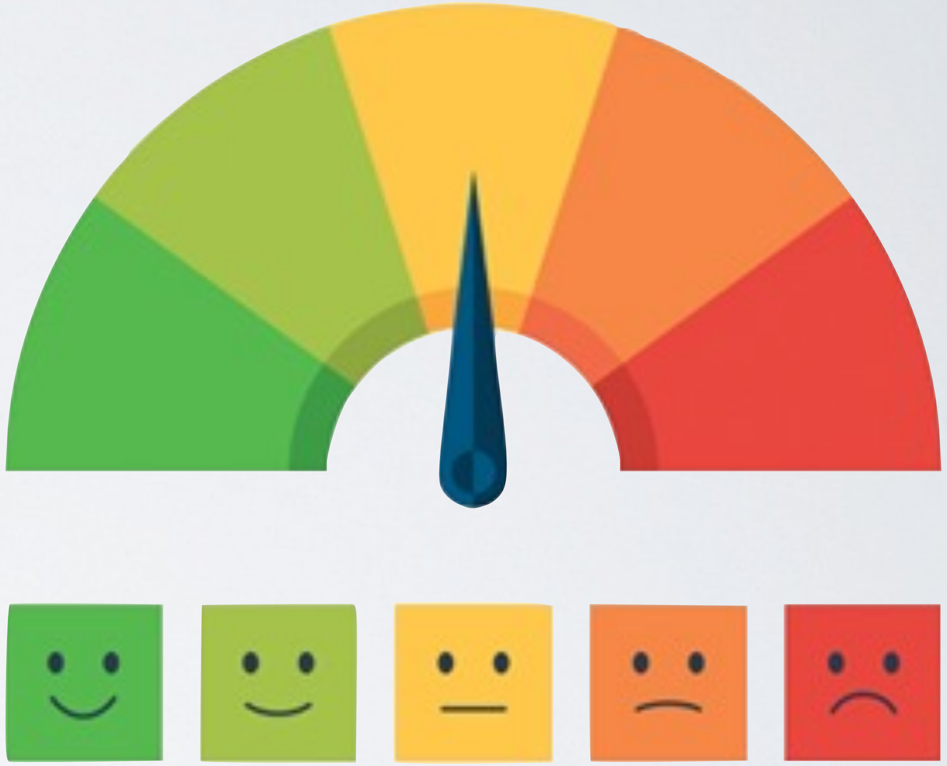
Lord William Thompson,
first Baron Kelvin (1824-1907)

What to Evaluate?

Efficiency

VS

Effectiveness



Critical Issues in Evaluation

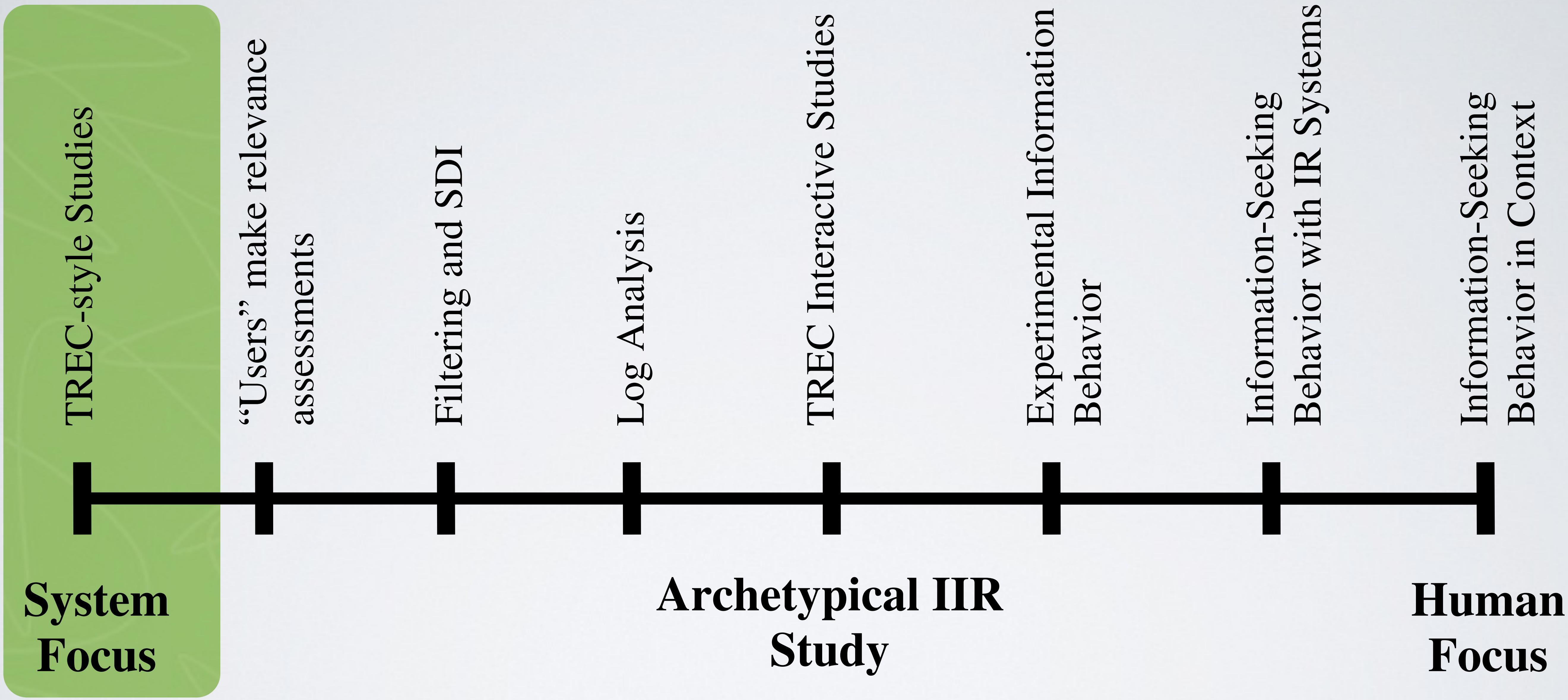
- It must be scientifically **valid**
  - valid methodology, measures, and statistics
  - large-scale enough to be statistically valid
  - must be "repeatable" if possible

- It must be **realistic** for the applications that will be using the information retrieval systems
  - **task** and use cases

- It must be **understandable** to your audience/client

Harman, D. K. (2011). *Information Retrieval Evaluation*. Morgan & Claypool Publishers, USA.

Evaluation Spectrum



Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. Foundations and Trends in Information Retrieval (FnTIR), 3(1-2), 1-224.

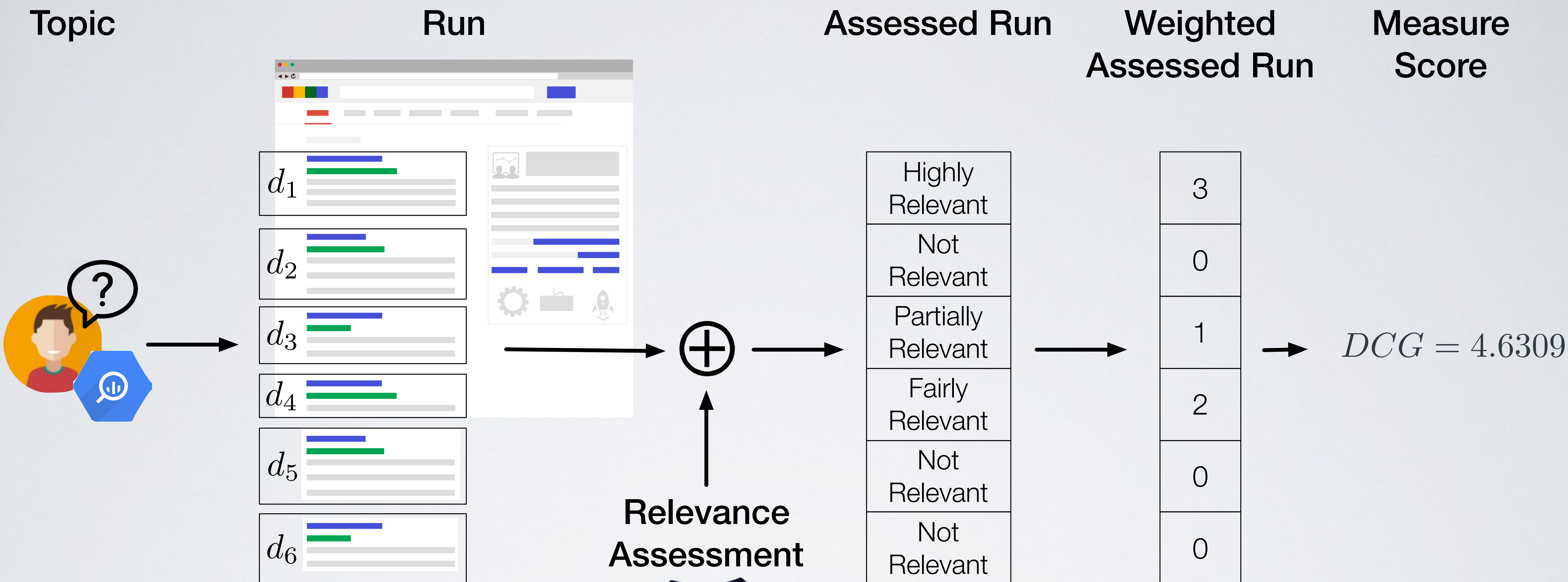
How Does Experimental Evaluation Work

- **Cranfield Paradigm** by Cyril W. Cleverdon
  - Dates back to mid 1960s
- Makes use of **experimental collections**
  - **documents** (corpora)
  - **topics**, which are a surrogate for information needs
  - **relevance judgments** (binary or graded) also called relevance assessment or ground-truth (or qrels)
- Ensures **comparability** and **repeatability** of the experiments

Cyril W. Cleverdon

Cleverdon, C. W. (1962). *Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems*. Aslib Cranfield Research Project, College of Aeronautics, Cranfield, UK.
Cleverdon, C. W. (1997). *The Cranfield Tests on Index Languages Devices. In Spärck Jones, K. and Willett, P., editors, Readings in Information Retrieval*, pages 47–60. Morgan Kaufmann Publisher, Inc., San Francisco, CA, USA.

Evaluation with Test Collections in a Nutshell

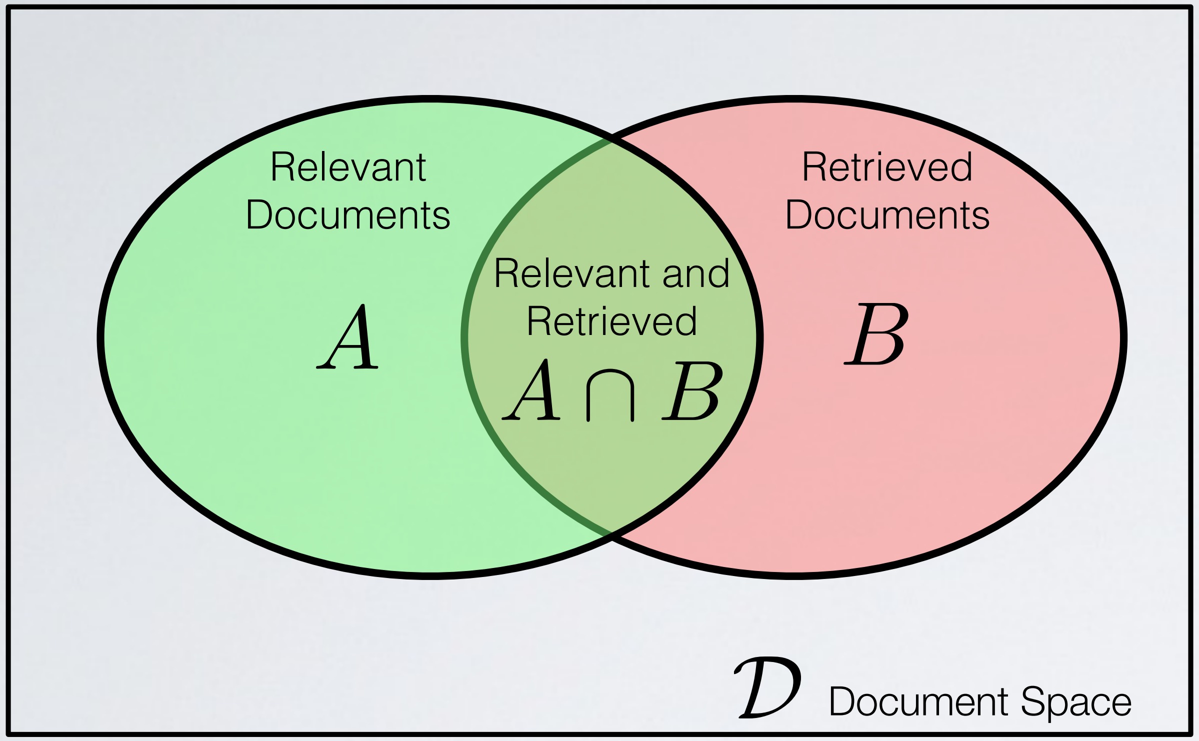


- Since we use set of topics, we can average the performance of a system over them
- We can compare two systems A and B run on the same test collection by comparing their average performance or,much better, by using statistical significance tests

Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval (FnTIR),* 4(4):247–375.

A Taxonomy of Evaluation Measures

	Set-Based Retrieval	Rank-Based Retrieval
Binary Relevance	Precision (P) Recall (R) F-measure (F)	Precision at Document Cut-off (P@k) Recall at Document Cut-off (R@k) R-Precision (Rprec) Average Precision (AP) Rank-Biased Precision (RBP) …
Multi-graded Relevance	Not widely agreed generalizations of Precision and Recall	Discounted Cumulated Gain (DCG) …



$$P = \frac{|A \cap B|}{|B|} \qquad R = \frac{|A \cap B|}{|A|}$$

$$F = \frac{2}{\frac{1}{P} + \frac{1}{R}} = 2\frac{P \cdot R}{P + R}$$

- **Precision** is the the proportion of retrieved documents that are actually relevant
- **Recall** is the the proportion of relevant documents actually retrieved
- Together, Precision and Recall measure **retrieval effectiveness**, meant as the ability of a system to retrieve relevant documents while at the same time holding back non-relevant ones
  - maximizing Precision and Recall corresponds to optimal retrieval in the sense of the **Probability Ranking Principle**, i.e. ordering documents by their decreasing probability of being relevant, and creates a tight link between retrieval models and evaluation
- **F-measure** is the harmonic mean of Precision and Recall, summarising them into a single score

van Rijsbergen, C. J. (1974). Foundations of Evaluation. *Journal of Documentation*, 30(4):365–373.
van Rijsbergen, C. J. (1981). Retrieval effectiveness. In Spärck Jones, K., editor, *Information Retrieval Experiment*, pages 32–43. Butterworths, London, United Kingdom.

Rank-based Measures: Average Precision

$$AP = \frac{1}{RB}\sum_{k \in \mathcal{R}} P(k) = \frac{1}{RB}\sum_{n=1}^{N} \underbrace{\left(\frac{1}{n}\sum_{m=1}^{n} r_m\right)}_{P(n)} r_n =$$

$$= \underbrace{\frac{rr}{RB}}_{\text{Recall}} \cdot \underbrace{\frac{1}{rr}\sum_{k \in \mathcal{R}} P(k)}_{\text{arithmetic mean of } P(k)}$$

where

- $\mathcal{R}$ is the set of the rank positions of the relevant retrieved documents
- $rr = |\mathcal{R}|$ is the total number of relevant retrieved documents
- N is the total number of retrieved documents, i.e. the length of the run
- The **Mean Average Precision (MAP)** is the mean of AP over a set of topics
  - Differently from the other measures, this mean has its own name since it is the most widely used single number to summarise the whole performance of a system

Chris Buckley

Buckley, C. and Voorhees, E. M. (2005). Retrieval System Evaluation. In Harman, D. K. and Voorhees, E. M., editors, *TREC. Experiment and Evaluation in Information Retrieval*, pages 53–78. MIT Press, Cambridge (MA), USA.

Rank-based Measures: Discounted Cumulated Gain

$$DCG(k) = \begin{cases} \sum_{n=1}^{k} r_n & \text{if } k < b \\ DCG(k-1) + \frac{r_k}{\log_b(k)} & \text{if } k >= b \end{cases} =$$

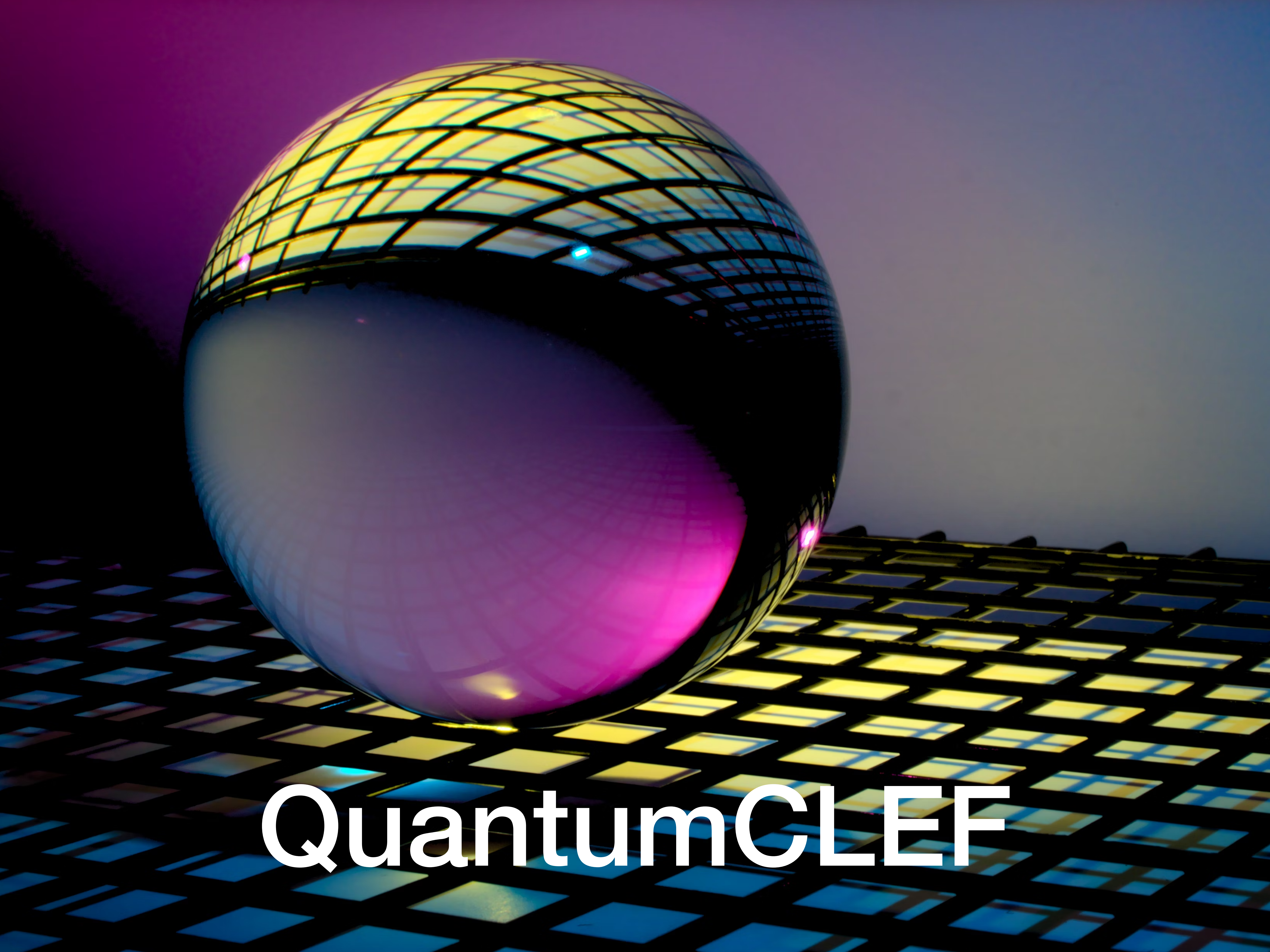
$$= \sum_{n=1}^{k} \frac{r_n}{\max(1, \log_b(n))}$$

- where the base of the logarithm b indicates the patience of the user in scanning the result list
  - $b = 2$ is an impatient user
  - $b = 10$ is a patient user
- DCG naturally handles multi-graded relevance
- DCG does not depend on the recall base
- DCG is not bounded in [0, 1]

Kalervo Järvelin

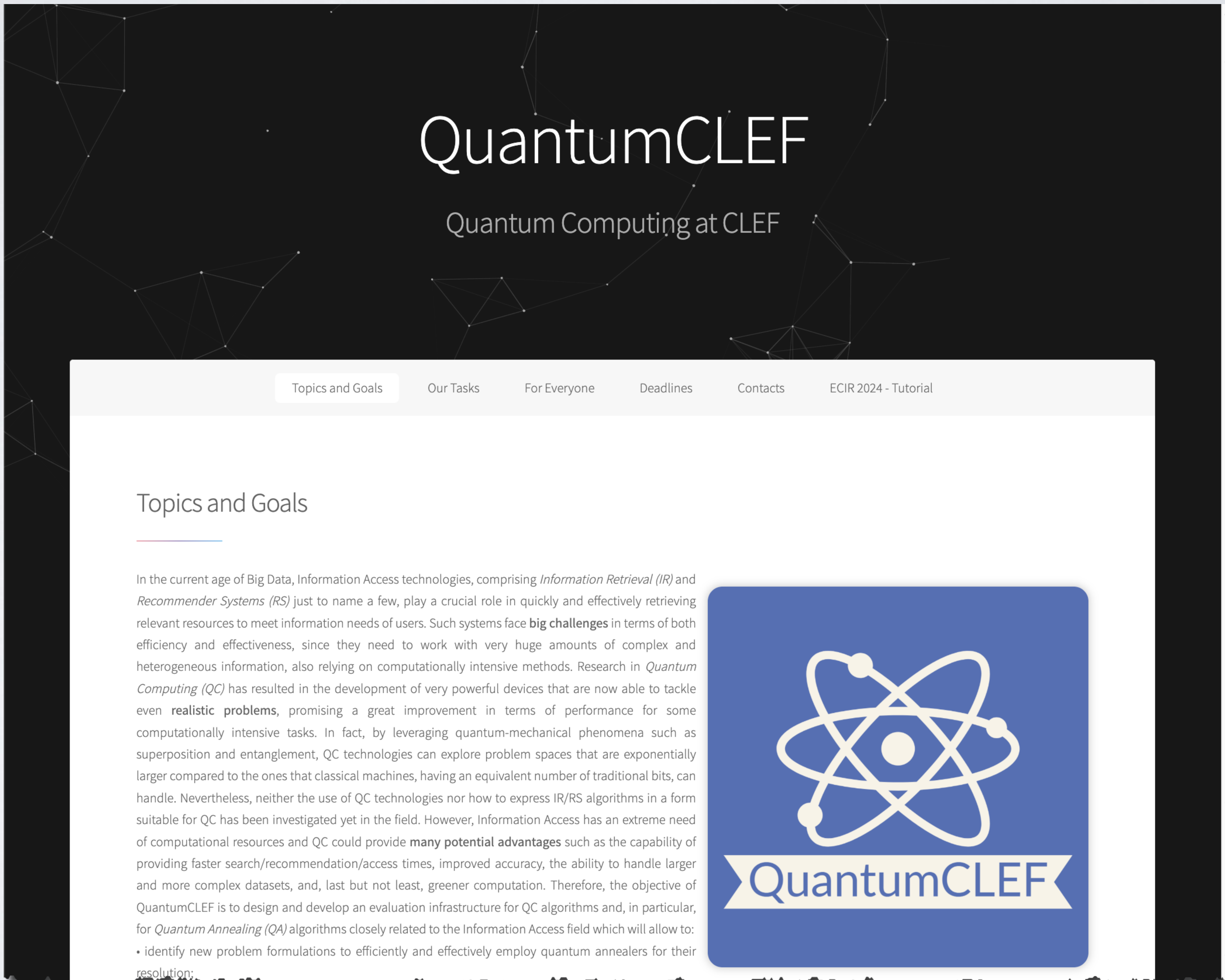
Jaana Kekäläinen

Järvelin, K. and Kekäläinen, J. (2002). Cumulated Gain-Based Evaluation of IR Techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446


QuantumCLEF

QuantumCLEF

https://qclef.dei.unipd.it/


QuantumCLEF

Quantum Computing at CLEF

Topics and Goals | Our Tasks | For Everyone | Deadlines | Contacts | ECIR 2024 - Tutorial

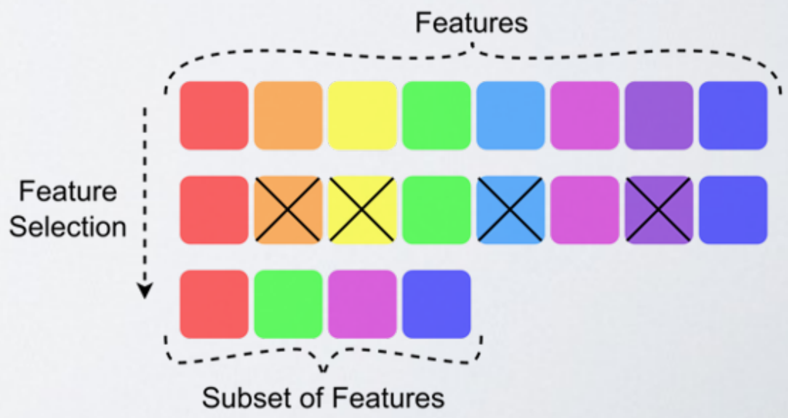
Topics and Goals

In the current age of Big Data, Information Access technologies, comprising *Information Retrieval (IR)* and *Recommender Systems (RS)* just to name a few, play a crucial role in quickly and effectively retrieving relevant resources to meet information needs of users. Such systems face **big challenges** in terms of both efficiency and effectiveness, since they need to work with very huge amounts of complex and heterogeneous information, also relying on computationally intensive methods. Research in *Quantum Computing (QC)* has resulted in the development of very powerful devices that are now able to tackle even **realistic problems**, promising a great improvement in terms of performance for some computationally intensive tasks. In fact, by leveraging quantum-mechanical phenomena such as superposition and entanglement, QC technologies can explore problem spaces that are exponentially larger compared to the ones that classical machines, having an equivalent number of traditional bits, can handle. Nevertheless, neither the use of QC technologies nor how to express IR/RS algorithms in a form suitable for QC has been investigated yet in the field. However, Information Access has an extreme need of computational resources and QC could provide **many potential advantages** such as the capability of providing faster search/recommendation/access times, improved accuracy, the ability to handle larger and more complex datasets, and, last but not least, greener computation. Therefore, the objective of QuantumCLEF is to design and develop an evaluation infrastructure for QC algorithms and, in particular, for *Quantum Annealing (QA)* algorithms closely related to the Information Access field which will allow to:

• identify new problem formulations to efficiently and effectively employ quantum annealers for their resolution;

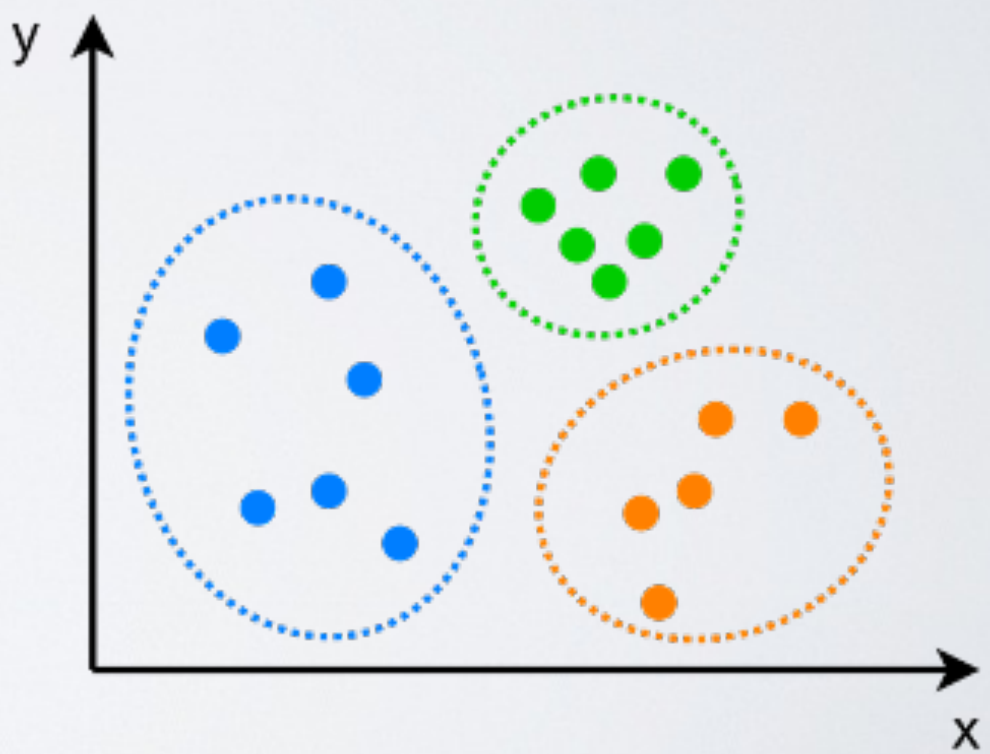
Task 1: Feature Selection

- **Task 1A - IR**: Select the most relevant features in the considered datasets to train a LambdaMART
  - A baseline is computed using Recursive Feature Elimination (RFE) with the Logistic Regression classifier
  - Datasets: MQ2007/LETOR and iSTELLA (challenging because of number of features)
  - Measure: nDCG@10
- **Task 1B - RS**: The task is to select the subset of features that will produce the best recommendation quality when used for an Item-Based KNN recommendation model
  - A baseline is computed using features selected by a bayesian search optimizing the model recommendation effectiveness
  - Datasets: 150_ICM and 500_ICM (music recommendation)
  - Measure: nDCG@10

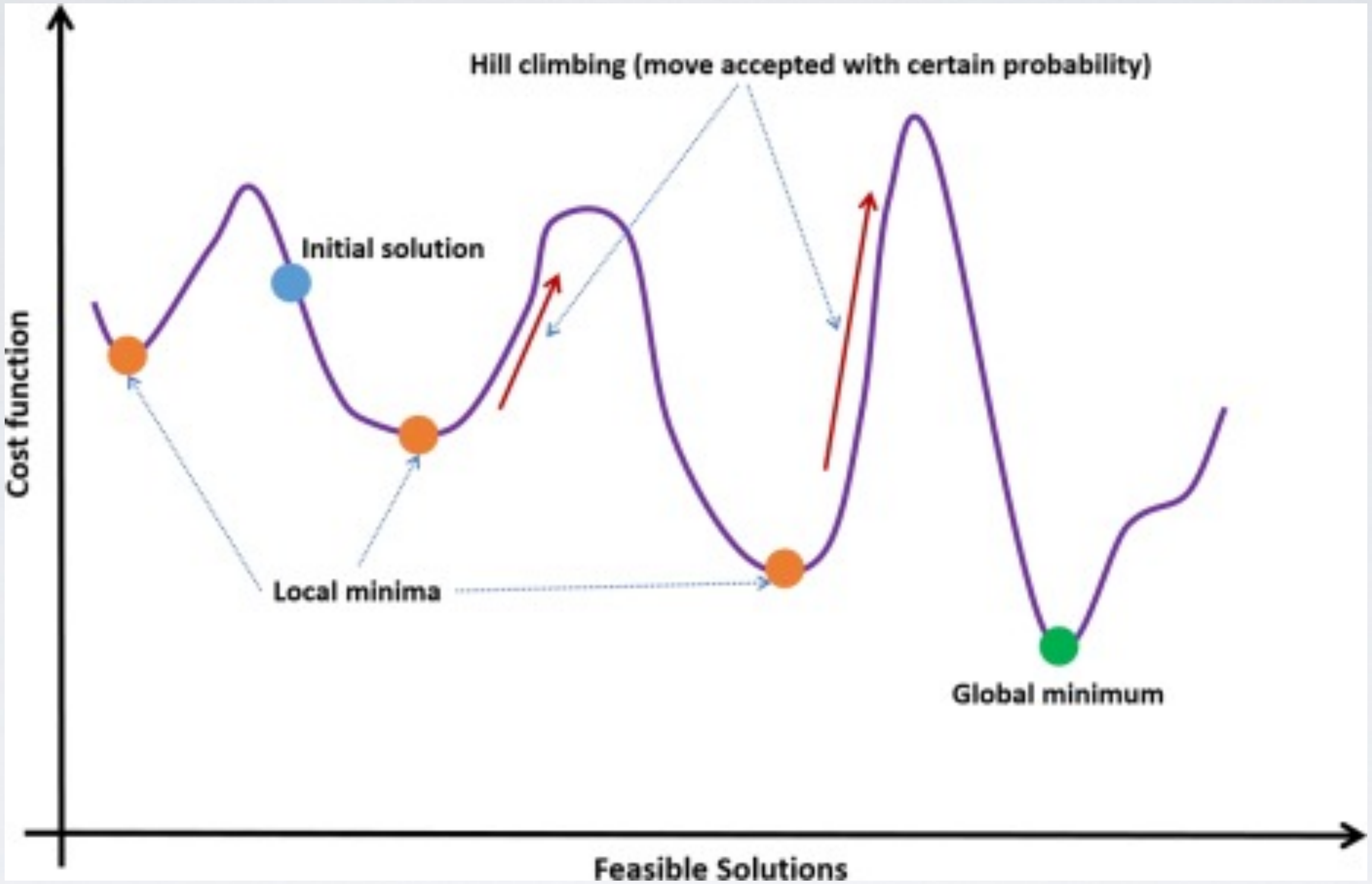


Task 2: Clustering

- Use QA to cluster different documents in the form of embeddings to ease processing of large collections
  - Obtain a list of representative centroids of the given dataset of embeddings
  - Search will be restricted to the clusters that are most likely to contain relevant documents, thereby reducing the search space and improving retrieval speed
- Dataset: ANTIQUE dataset
- Measure:
  - the Davies-Bouldin Index will be used to measure the overall cluster quality
  - the nDCG@10 will be used to measure the overall retrieval quality



Simulated Annealing



- For both tasks (Feature Selection and Clustering), the QA approach will be compared against a SA approach, using the same QUBO formulation

Effectiveness Challenges

Quantum Computing for Information Access - Feature Selection.ipynb

File Edit View Insert Runtime Tools Help Last edited on November 20

Comment Share

+ Code + Text Connect

Now we will use Quantum Annealing for real. This works as follows:

1. We need to have an API KEY which is required to have access to D-Wave's infrastructure.
2. Once we create our sampler and submit our problem, we will send it through internet. The problem will reach the D-Wave's infrastructure and will be enqued if there are currently other problems running.
3. Once it is our turn, our problem will be solved and the solution will be sent back to us.

```
[ ] %%time

    from google.colab import userdata
    token=userdata.get('API_TOKEN')

    sampler_QPU = DWaveCliqueSampler(token=token)

    response_QPU_4 = sampler_QPU.sample(kbqm,
                            label='Example - MI Feature Selection',
                            num_reads=num_reads)
```

```
CPU times: user 3.56 s, sys: 50.9 ms, total: 3.61 s
Wall time: 8.29 s
```

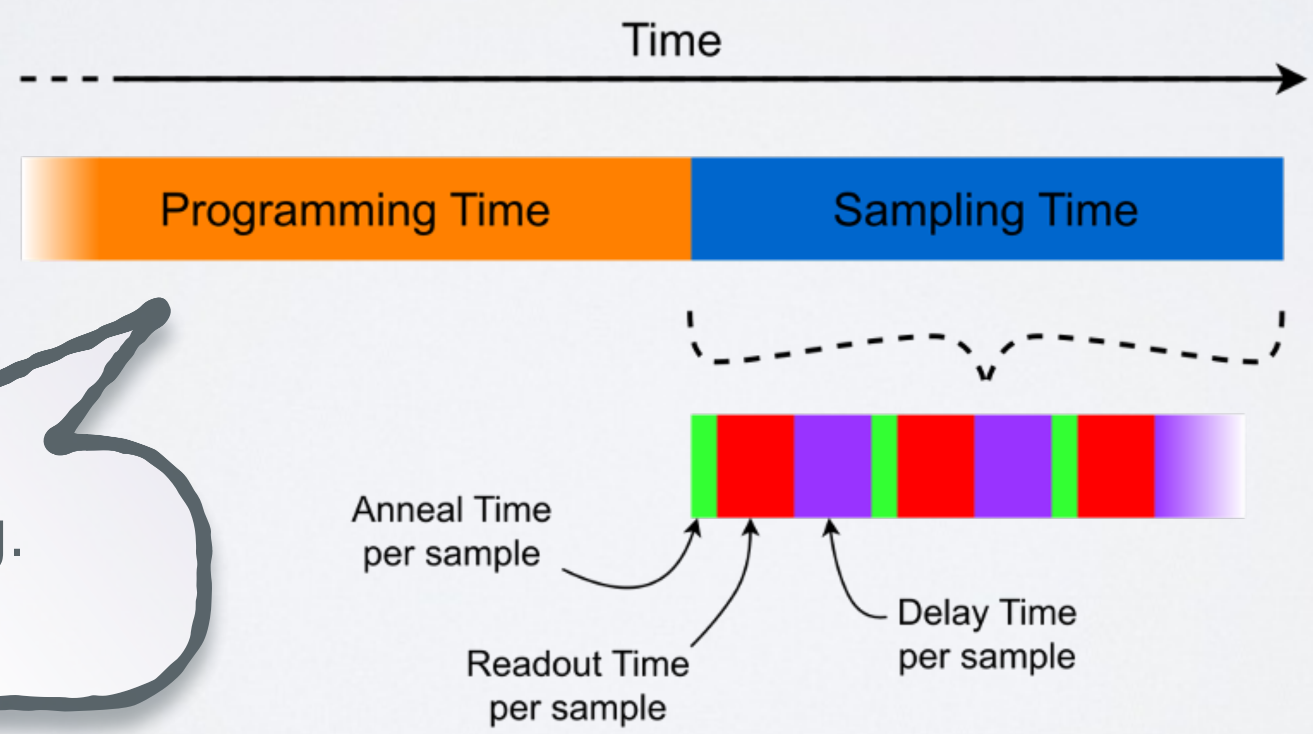
```
[ ] print_response_data(response_QPU_4.aggregate())
```

```
------------------------------------------------------------------
          Set 0           Set 1     Energy           Count
------------------------------------------------------------------
['age', 'alone', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'miss', 'mrs', 'rare']['cabin', 'mr', 'pclass'
['age', 'alone', 'cabin', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'miss', 'mrs']['mr', 'pclass', 'rare'
['age', 'alone', 'cabin', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'miss', 'rare']['mr', 'mrs', 'pclass'
['age', 'alone', 'cabin', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'mr', 'rare']['miss', 'mrs', 'pclass'
['age', 'alone', 'embarked port Q', 'embarked port S', 'fare', 'master', 'miss', 'mrs', 'pclass', 'rare']['cabin', 'embarked port C', 'mr'
['alone', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'miss', 'mrs', 'pclass', 'rare']['age', 'cabin', 'mr'
['age', 'alone', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'miss', 'mrs', 'pclass', 'rare']['cabin', 'master', 'mr'
['age', 'alone', 'embarked port C', 'embarked port Q', 'fare', 'master', 'miss', 'mrs', 'pclass', 'rare']['cabin', 'embarked port S', 'mr'
['age', 'alone', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'mrs', 'pclass', 'rare']['cabin', 'miss', 'mr'
['age', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'miss', 'mrs', 'pclass', 'rare']['alone', 'cabin', 'mr'
['age', 'alone', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'mr', 'pclass', 'rare']['cabin', 'miss', 'mrs'
['age', 'alone', 'embarked port C', 'embarked port Q', 'embarked port S', 'fare', 'master', 'miss', 'rare', 'sex']['cabin', 'mr', 'mrs', '
['age', 'alone', 'cabin', 'embarked port Q', 'embarked port S', 'fare', 'master', 'miss', 'pclass', 'rare']['embarked port C', 'mr', 'mrs'
```

Can we leverage the stochastic nature of the QA process to perform a deeper statistical analysis?

Efficiency Challenges

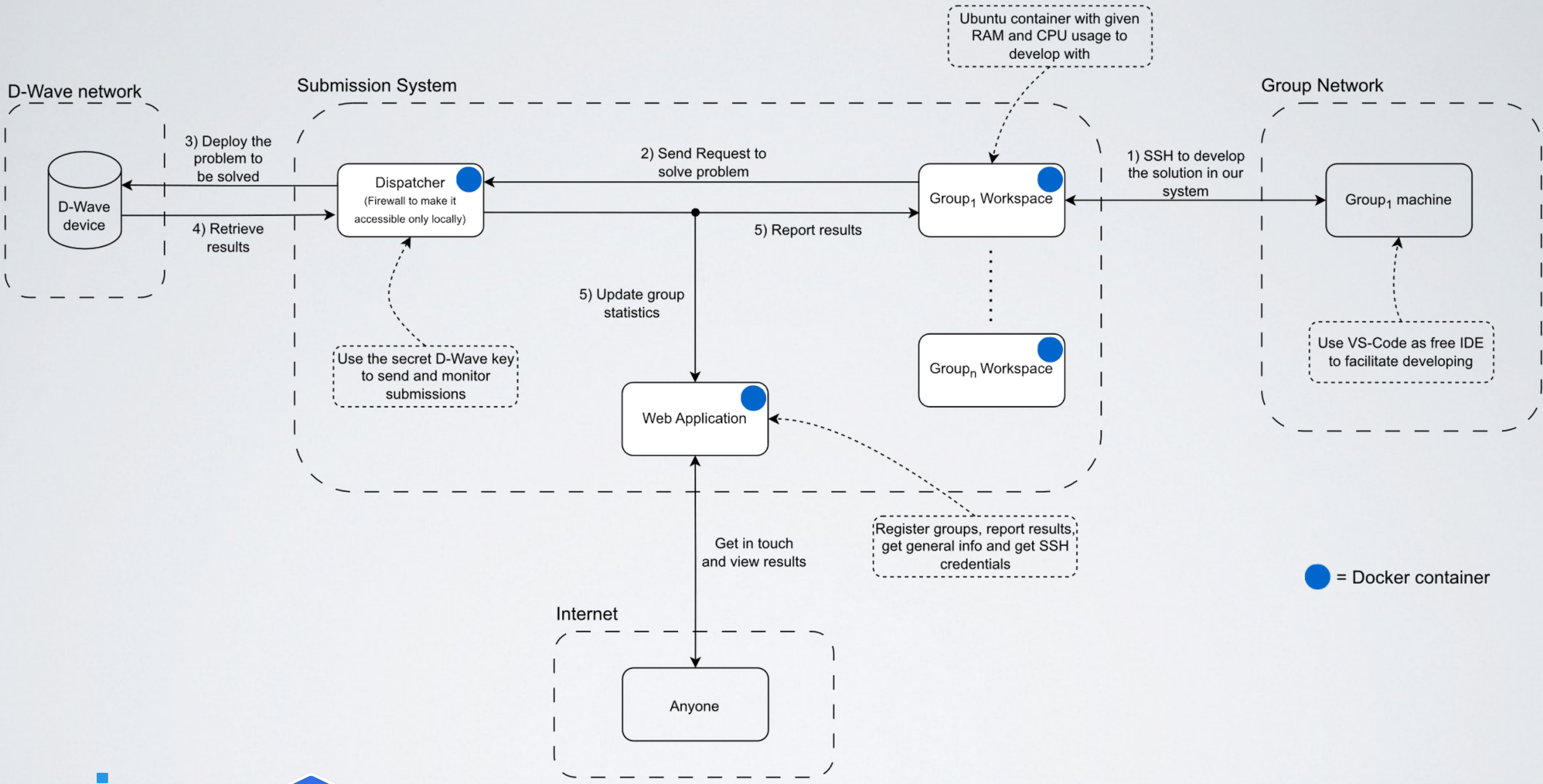
- There is not a standard way to measure the efficiency of Quantum Annealers
- There are several steps in the Annealing phase, each requiring a different amount of time based also on the used quantum annealer



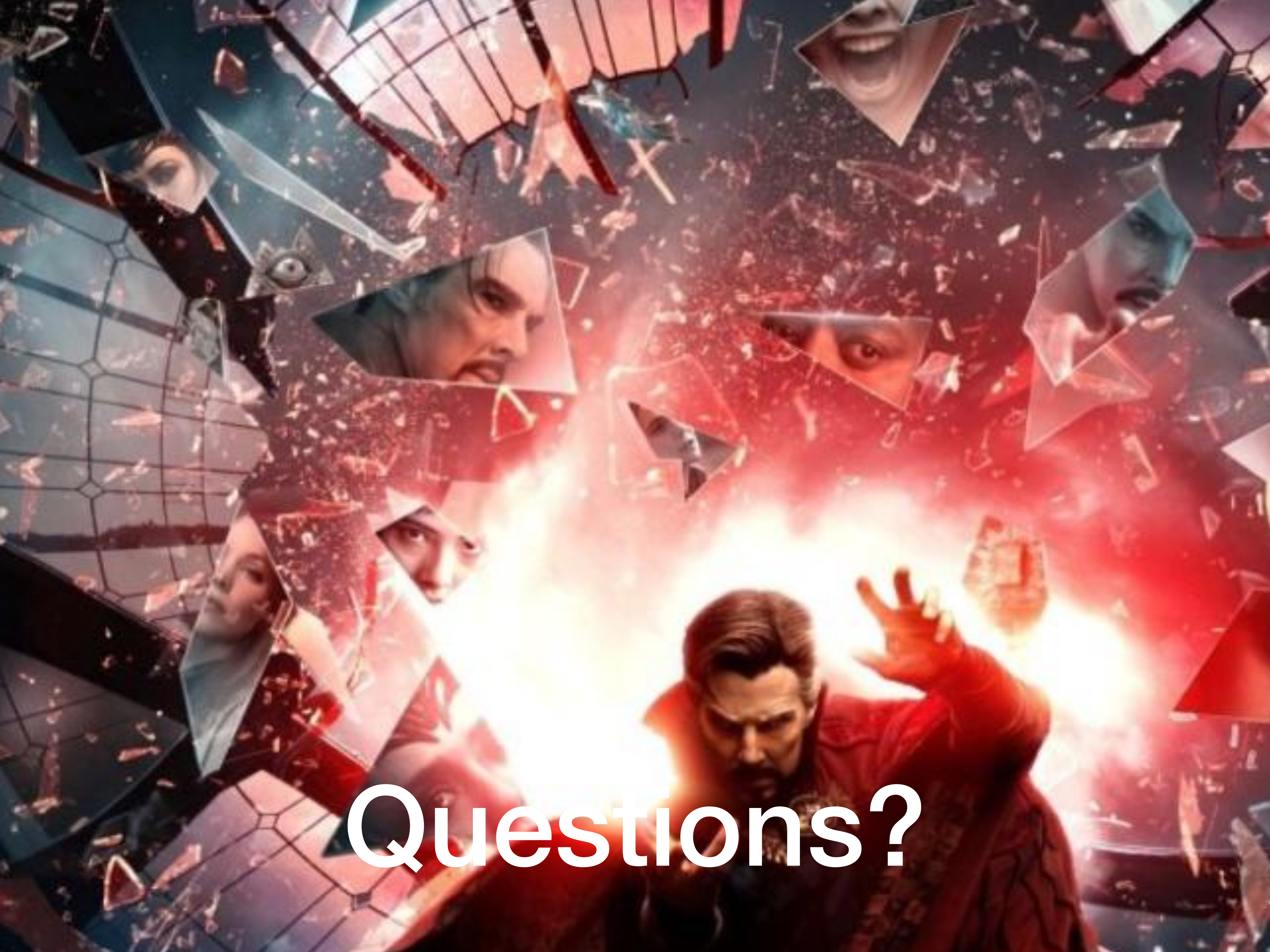
Minor embedding. Can we cache?

Resch Resch, S. and Karpuzcu, U. R. (2022). Benchmarking Quantum Computers and the Impact of Quantum Noise. *ACM Computing Surveys (CSUR)*, 54(7):142:1–142:35.

QuantumCLEF Infrastructure



docker kubernetes


Questions?