

# GSI024 - Organização e recuperação de informação

Prof. Dr. Rodrigo Sanches Miani (FACOM/UFU)

Última atualização - Junho/2022

# Avaliação da recuperação - Parte 1

# Agenda

## **“Avaliação da recuperação de informação”**

Introdução

Paradigma de Cranfield

Métricas de recuperação

Adequação das métricas

QP-2

# QP-2

- Marcado para o dia 30/06, próxima quinta-feira;
- Ponderação de termos, modelo vetorial e modelo probabilístico.

**Aula passada**

# Introdução

- O pré-processamento de documentos é um importante procedimento empregado na construção de sistemas de RI;
- Pode ser dividido em quatro operações (ou transformações) textuais:
  1. Análise léxica do texto (números, pontuação, caixa alta/baixa);
  2. Eliminação de stopwords;
  3. Stemming das palavras;
  4. Seleção de termos ou palavras-chave;

# Introdução



# Introdução

Para avaliar um sistema de RI, é necessário medir o quão bem o sistema atende a necessidade de informação do usuário;

- Isso pode ser um problema, pois um mesmo conjunto resposta pode ser interpretado de maneiras diferentes por usuários distintos;
- Mesmo assim, é possível definir métricas aproximadas, que, na média, têm uma correlação com as preferências de uma população.

# Motivação

Sem uma avaliação adequada, não temos como saber como o sistema de RI está desempenhando, nem podemos comparar objetivamente a qualidade e recuperação com a de outros sistemas.

- Uma modificação na função de ranqueamento é proposta: devemos ir adiante e implantá-la?
- Uma nova função de ranqueamento probabilística foi projetada: ela é superior ao ranqueamento do modelo vetorial?
- Para quais tipos de consultas à Web, como consultas sobre negócios, produtos ou consultas geográficas, uma modificação no ranking funciona melhor?

# Definição

- ***Avaliação da recuperação** é um processo sistemático no qual se associa uma métrica quantitativa aos resultados produzidos por um sistema de RI em resposta a um conjunto de consultas de usuário.*
- Essa métrica deve ser diretamente associada à **relevância** dos resultados para os usuários.
- Uma abordagem comum para computar tal métrica é comparar o resultado produzido pelo sistema com os resultados sugeridos por humanos para o mesmo conjunto de consultas.

# Importante!

- Avaliação da recuperação significa avaliar a **qualidade** dos resultados, não o desempenho do sistema;
- A definição anterior não cobre aspectos que afetam o julgamento do usuário:
  - Interface do usuário;
  - Contexto da busca;
  - Período em que a busca foi feita.

# Apesar das deficiências...

O processo de associar uma métrica aos resultados da consulta continua sendo amplamente utilizado:

- Simplicidade;
- Repetibilidade;
- Custos relativamente baixos.

# Paradigma de Cranfield

# Breve histórico – Cranfield 1

- Em 1952, Cyril Cleverdon, bibliotecário da escola de aeronáutica de Cranfield, Inglaterra, conheceu o sistema Uniterm proposto por Mortimer Taube, bibliotecário nos EUA;
- Uniterm – 40 mil títulos formados por 7 mil palavras distintas;
- Cleverdon achou o sistema interessante, indexou 200 documentos manualmente usando o sistema Uniterm e pediu para um colega que fizesse diversas consultas.

# Breve histórico – Cranfield 1

- Uniterm x Sistemas de indexação mais sofisticados: briga entre os defensores de cada um deles;
- Cleverdon propôs um estudo para comparar os diversos sistemas de indexação existentes;
- Cada consulta era baseada em um só documento e a busca era considerada bem-sucedida se aquele documento fosse localizado no catálogo;



# Breve histórico – Cranfield 1

- Indexação manual de 18000 artigos sobre Engenharia Aeronáutica e avaliação de 1200 consultas;
- Os resultados mostraram que os quatro sistemas de indexação eram basicamente equivalentes em termos da **precisão** dos resultados;
  - Fração dos documentos recuperados que é de fato relevante.
- Os resultados também mostraram que a **revocação (recuperação)** era pouco útil se não estivesse acompanhada sobre a **precisão**;
  - Revocação: fração dos documentos relevantes que foi recuperada.
- Dentro de um sistema de RI não era possível aumentar simultaneamente a precisão e a revocação.

# Breve histórico – Cranfield 2

- 1400 documentos e 279 consultas;
- Para cada consulta, todos os documentos da coleção foram examinados e sua relevância em relação à consulta foi determinada;
- Seis estudantes passaram três meses examinando cada documento em relação a consulta e decidindo quais documentos eram relevantes;
- O resultado foi uma **coleção de referência**.

# Breve histórico – Cranfield 2

- Em situações práticas, a maioria das buscas não requer uma revocação (recuperação) alta;
- A maioria dos usuários requer apenas algumas respostas relevantes (precisão);
- Os experimentos Cranfield 2 estabeleceram a base para a **experimentação moderna em RI**;
  - O mesmo conjunto de documentos e consultas pode ser usado para avaliar sistemas de ranqueamento diferentes comparando-os com os julgamentos de relevância produzidos por especialistas humanos.

# Coleção de referência

- Coleções de referência permitem comparar diretamente os resultados produzidos por diferentes funções de ranqueamento;
- Os julgamentos de relevância são produzidos por humanos especialistas e idealmente devem fornecer uma decisão de relevância para cada par necessidade de informação-documento;
- Claramente, isso só é viável para coleções de documento pequenas, como as dos experimentos Cranfield.

# Coleção de referência - Vantagens

- Dada uma coleção de referência, uma avaliação do sistema de RI pode ser feita rapidamente;
- Avaliações de sistemas de RI podem ser reproduzidas posteriormente para fins de verificação (Repetibilidade);
- Coleções de referência diferentes podem ser construídas focando em tipos particulares de necessidade de informação.

# Métricas de recuperação

# Precisão e revocação

Considere uma requisição de informação  $I$  (de uma coleção de referência) e seu conjunto  $R$  de relevantes. Seja  $|R|$  o número de documentos nesse conjunto.

Suponha que um dado algoritmo de recuperação (que está sendo avaliado) processa a requisição de informação  $I$  e gera um conjunto resposta  $A$ . Seja  $|A|$  o número de documentos nesse conjunto e  $|R \cap A|$  o número de documentos na intersecção dos conjuntos  $R$  e  $A$ .

# Precisão e revocação

As medidas de precisão e revocação são definidas da seguinte forma:

**Precisão** (fração dos documentos recuperados que é relevante):

$$p = |R \cap A| / |R|$$

**Revocação** (fração dos documentos relevantes que foi recuperada):

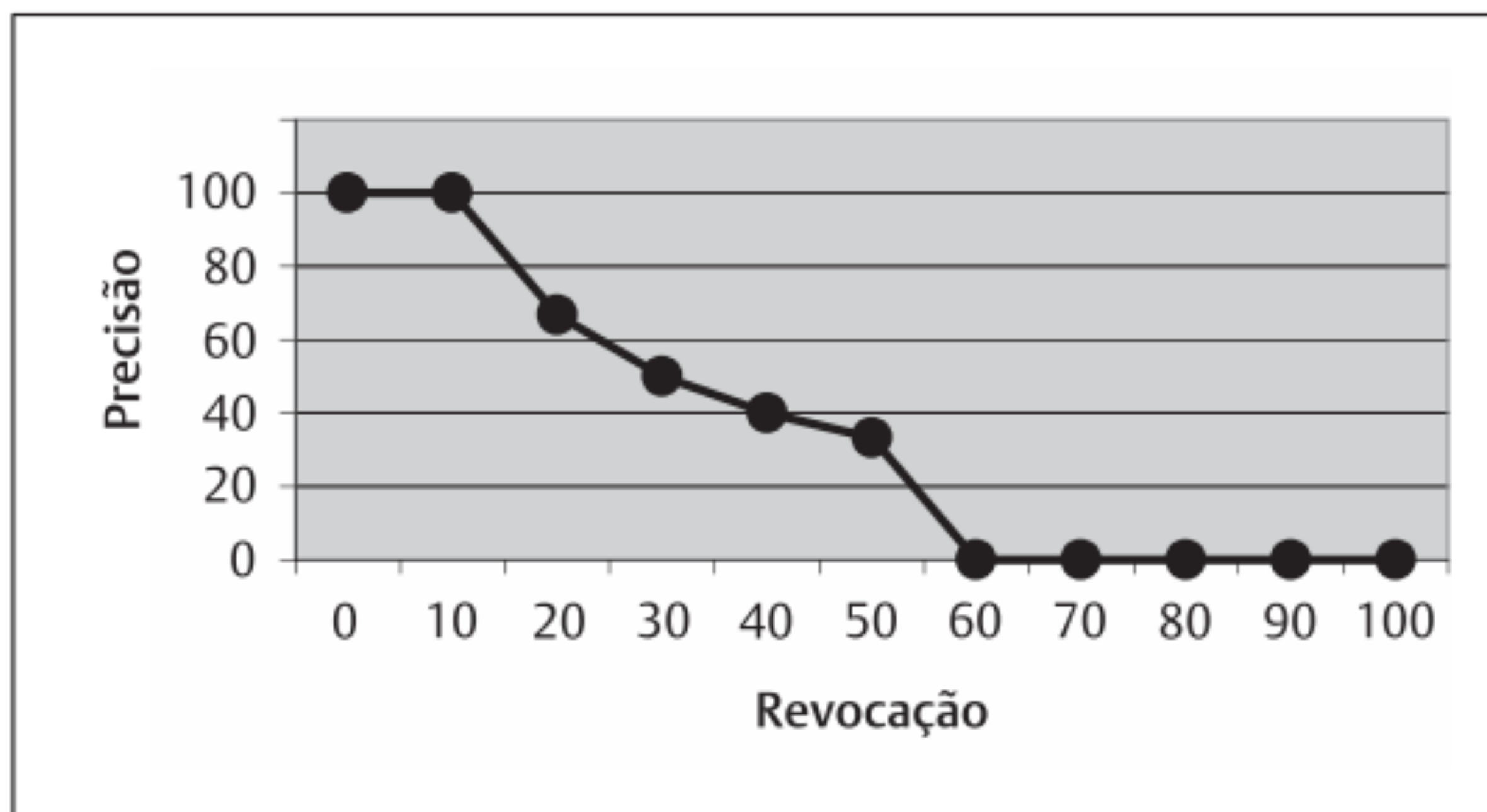
$$r = |R \cap A| / |A|$$



# Precisão e revocação

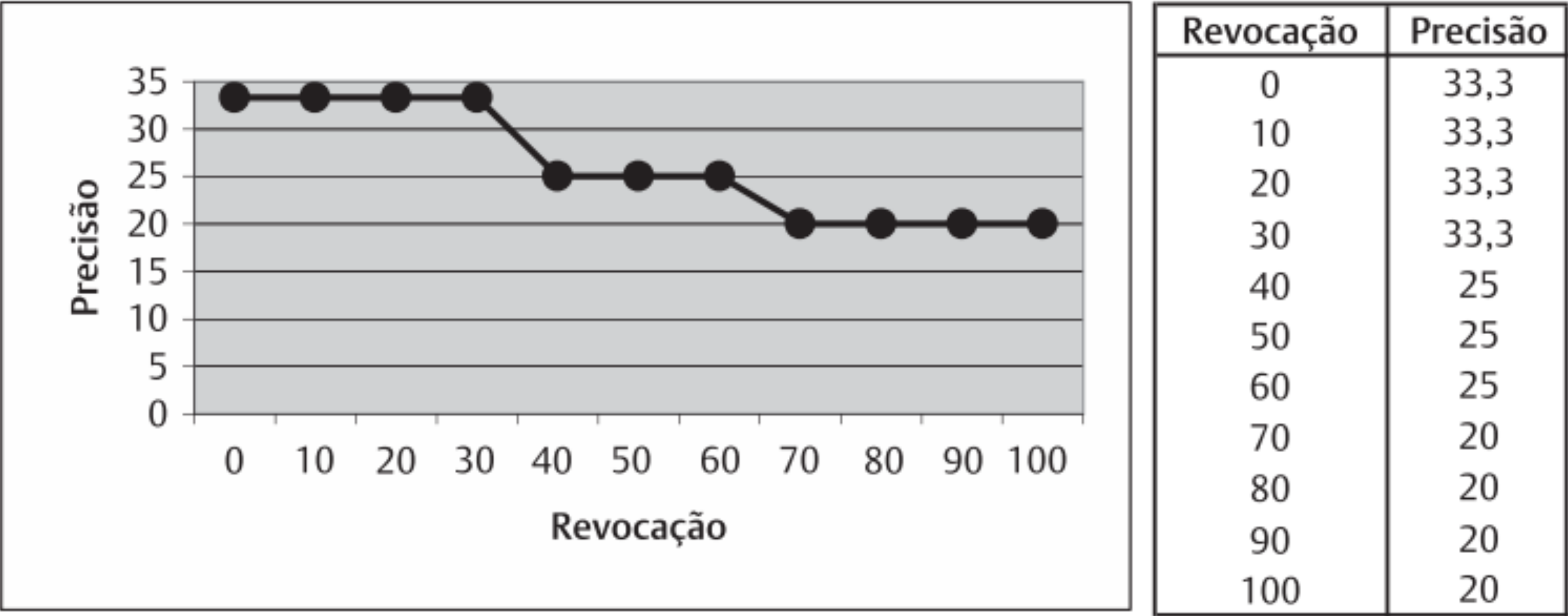
- Tais métricas supõem que todos os documentos no conjunto resposta A foram examinados;
- Contudo, o usuário não é normalmente apresentado a todos os documentos do conjunto resposta A de uma vez;
- Em vez disso, os documentos do conjunto A são primeiro ordenados de acordo com uma função de ranqueamento;
- O usuário examina essa lista ordenada iniciando pelo documento do topo:
  - Nesse caso, as medidas de precisão e revocação variam conforme o usuário procede com seu exame do conjunto resposta.

# Precisão e revocação – Exemplo – q1

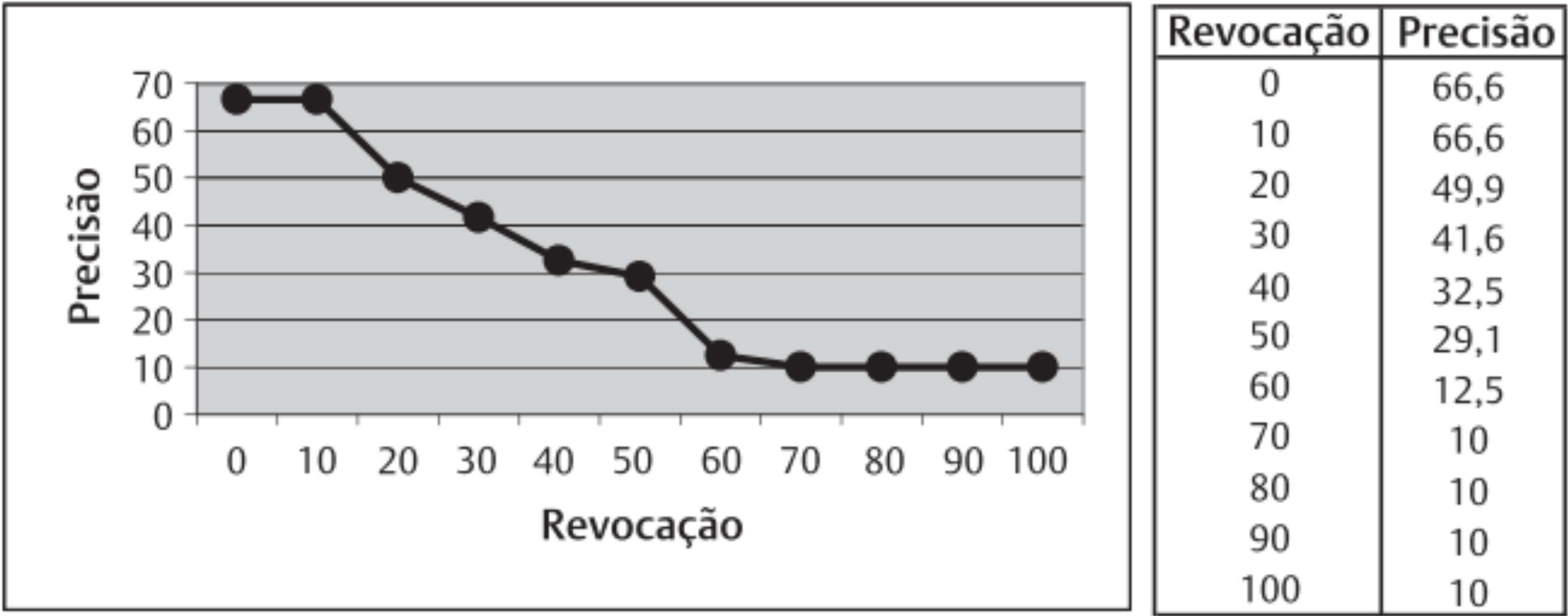


Revocação	Precisão
0	100
10	100
20	66,6
30	50
40	40
50	33,3
60	0
70	0
80	0
90	0
100	0

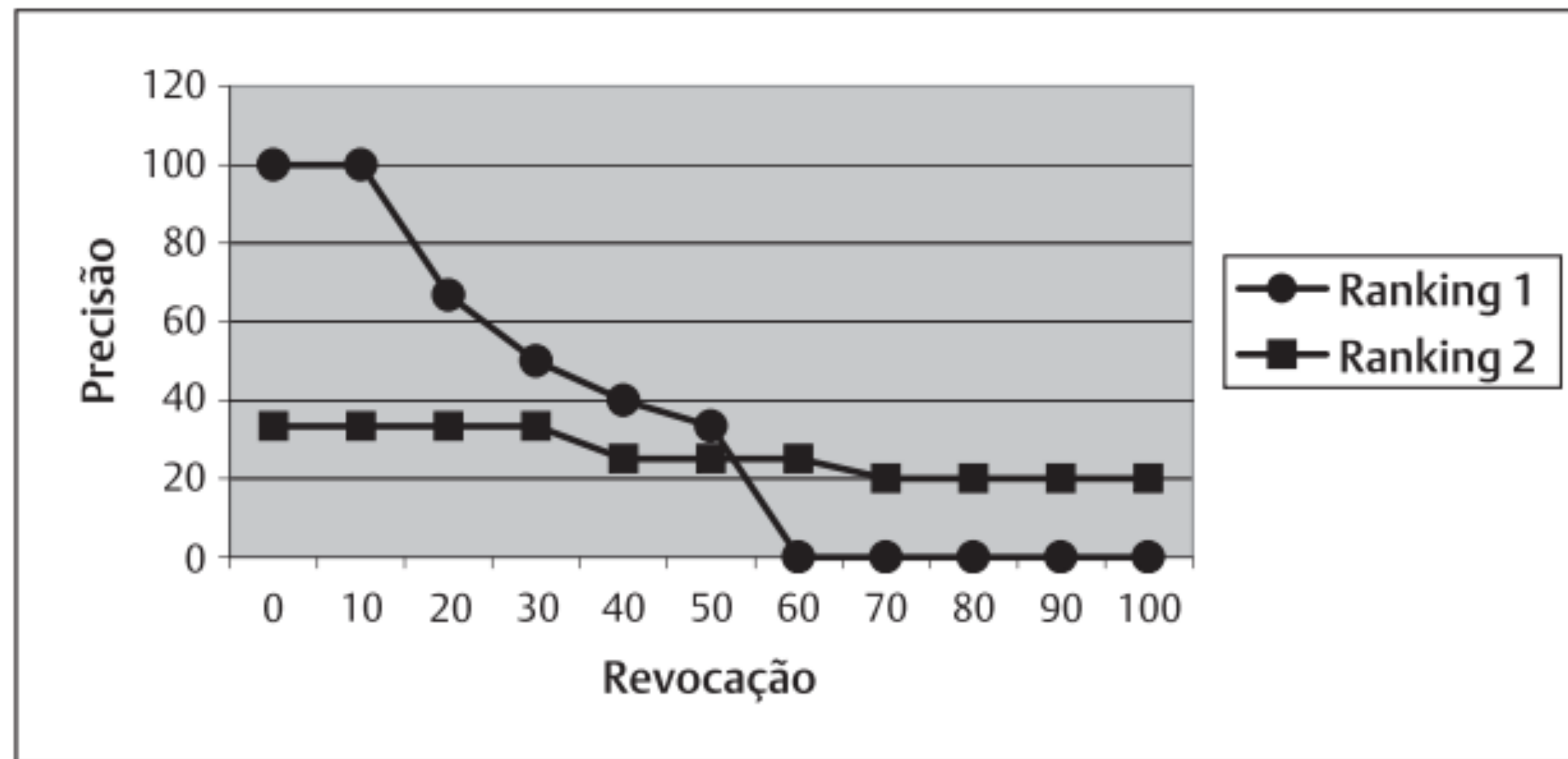
# Precisão e revocação – Exemplo – q2



# Precisão e revocação – Exemplo (média entre q1 e q2)



## Precisão e revocação – Exemplo (quaisquer dois sistemas de RI )



# Adequação das métricas

# Precisão e revocação - Problemas

- Precisão e revocação têm sido amplamente utilizadas para avaliar a qualidade de algoritmos de recuperação;
- Uma reflexão mais cuidadosa revela problemas com essas medidas;
- Veremos cada um deles.

# Precisão e revocação - Problemas

1. A estimativa da **revocação máxima** para uma consulta requer um conhecimento detalhado de todos os documentos da coleção – para coleções grandes tal conhecimento não está disponível;
2. Precisão e revocação são medidas relacionadas que capturam diferentes aspectos do conjunto de documentos recuperados – em certos casos uma só medida pode ser mais apropriada;
3. Precisão e revocação medem a eficácia para um conjunto de consultas processadas em lotes – sistemas modernos a interatividade é um aspecto fundamental do processo de recuperação.



# Comentários

# No decorrer da aula vimos...

- Como avaliar um modelo de recuperação da informação usando uma coleção de referência;
- Duas métricas amplamente utilizadas foram apresentadas: precisão e revocação.
- **Precisão** – representa a fração de documentos recuperados que são relevantes;
- **Revocação** – representa a fração de documentos relevantes que foram recuperados.

# No decorrer da aula vimos...

- Algoritmos de RI voltados para a busca Web devem apresentar níveis de precisão mais altos;
- Algoritmos de RI para aplicações específicas como as da área da saúde ou jurídica necessitam de níveis altos de revocação;
- Valores de precisão versus revocação são atualmente medidas de avaliação padrão para sistemas de RI e usadas extensivamente na literatura.

# Próximas aulas – Avaliação da recuperação

- Sumários com um único valor;
- Comparação da ordenação relativa produzida por dois algoritmos – Correlação;
- Atividade.