



SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

CONCEPTOS BÁSICOS SOBRE MEDIDAS DE RECUPERACIÓN

[Objetivo]

Evaluar dos buscadores por ejemplo Google y Yahoo Search respecto a la precisión y a la exhaustividad

[Método]

1. Realizar sin comillas la siguiente consulta en los buscadores:
 - a) *Las 7 maravillas del mundo*
 - b) *Como posicionar bien los resultados de una consulta en Internet*
2. Evaluar entre los 15 primeros resultados recuperados no patrocinados que resultados son relevantes, en cada buscador y para cada una de las preguntas. Se considera un documento relevante si determina y satisface la necesidad de información en el documento o como máximo accediendo a esa información a través de un enlace del resultado. En el caso de:
 - a) se considerará relevante cuando aparezcan las originarias 7 maravillas del mundo, y no si únicamente aparecen las nuevas.
 - b) deben darse directrices de cómo posicionar bien los resultados.
3. Calcular la precisión y la recall para cada una de las consultas y para cada buscador. A efectos de esta actividad se estima que para esa consulta existen 10 documentos relevantes en la colección (Internet).
4. Hallar la media de precisión de cada buscador para los once niveles de recall. Representar la precision-recall de cada buscador en un gráfico.
5. Representar en una única gráfica precisión-recall ambos buscadores y comentar los resultados.
6. Realizar un histograma de rprecision para cada consulta. Comentar los resultados.

[Herramientas]

El uso de herramientas como Microsoft Excel puede facilitar la tarea de los cálculos y dibujar las gráficas.





SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

T relevantes: 10
T recuperado: 15

| RECUPERADO | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| Google | r | r | r | r | n | n | r | n | r | n | n | r | r | n | n |
| Yahoo | r | r | n | r | r | r | r | n | r | n | n | n | r | r | r |
| Google | 1 | 2 | 3 | 4 | 4 | 4 | 5 | 5 | 6 | 6 | 6 | 7 | 8 | 8 | 8 |
| Yahoo | 1 | 2 | 2 | 3 | 4 | 5 | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 10 |

Google

| | | | | | | | | | | | | | | | |
|-----------|------|------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Recall | 10% | 20% | 30% | 40% | 40% | 40% | 50% | 50% | 60% | 60% | 60% | 70% | 80% | 80% | 80% |
| Precision | 100% | 100% | 100% | 100% | 80% | 67% | 71% | 63% | 67% | 60% | 55% | 58% | 62% | 57% | 53% |

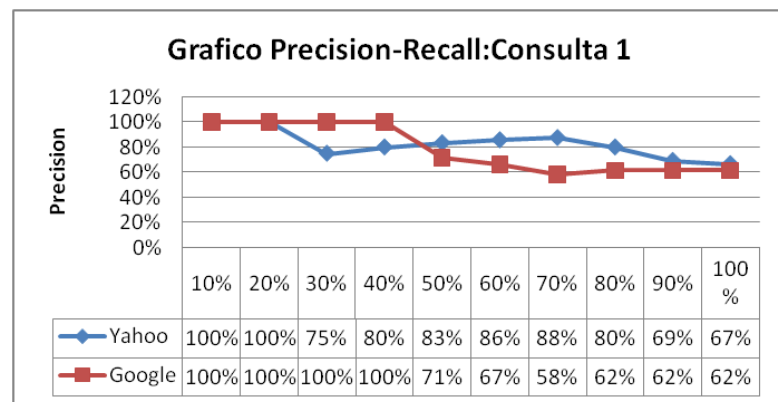
Yahoo

| | | | | | | | | | | | | | | | |
|-----------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| Recall | 10% | 20% | 20% | 30% | 40% | 50% | 60% | 70% | 70% | 80% | 80% | 80% | 90% | 90% | 100% |
| Precision | 100% | 100% | 67% | 75% | 80% | 83% | 86% | 88% | 78% | 80% | 73% | 67% | 69% | 64% | 67% |

| | | | | | | | | | | | | | | |
|------|------|------|------|-----|-----|-----|-----|-----|-----|--|--|--|--|--|
| 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | | | | | | | |
| 100% | 100% | 100% | 100% | 71% | 67% | 58% | 62% | | | | | | | |
| 100% | 100% | 75% | 80% | 83% | 86% | 88% | 80% | 69% | 67% | | | | | |

| CONSULTA 1 | |
|--------------|-----|
| Media Google | 82% |
| Media Yahoo | 83% |

| | | | | | | | | | | |
|--------|------|------|------|------|-----|-----|-----|-----|-----|------|
| Yahoo | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| Google | 100% | 100% | 75% | 80% | 83% | 86% | 88% | 80% | 69% | 67% |
| Google | 100% | 100% | 100% | 100% | 71% | 67% | 58% | 62% | 62% | 62% |





SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

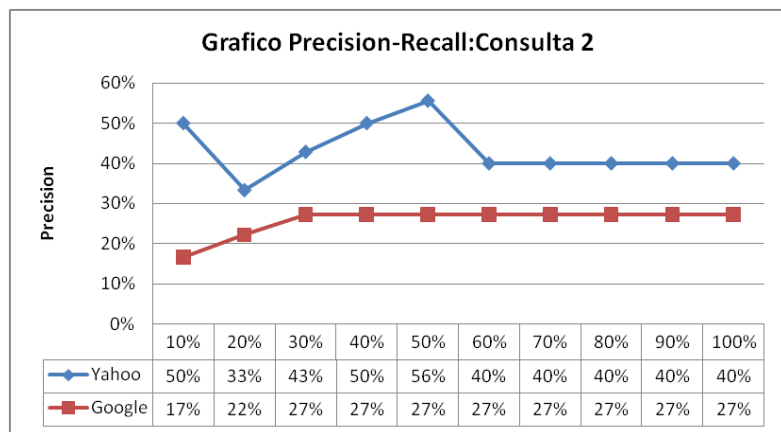
T relevantes: 10
T recuperado: 15

| RECUPERADO | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| Google | n | n | n | n | n | r | r | r | n | n | n | n | n | r | |
| Yahoo | n | n | n | n | r | n | n | r | n | r | n | n | n | n | |
| Google | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Yahoo | 0 | 1 | 1 | 1 | 1 | 2 | 3 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 6 |
| Yahoo | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |

| Google | 0% | 10% | 10% | 10% | 10% | 20% | 30% | 40% | 50% | 50% | 50% | 50% | 50% | 50% | 60% |
|-----------|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Recall | 0% | 10% | 10% | 10% | 10% | 20% | 30% | 40% | 50% | 50% | 50% | 50% | 50% | 50% | 60% |
| Precision | 0% | 50% | 33% | 25% | 20% | 33% | 43% | 50% | 56% | 50% | 45% | 42% | 38% | 36% | 40% |
| Yahoo | 0% | 0% | 0% | 0% | 0% | 10% | 10% | 10% | 20% | 20% | 30% | 30% | 30% | 30% | 30% |
| Recall | 0% | 0% | 0% | 0% | 0% | 10% | 10% | 10% | 20% | 30% | 30% | 30% | 30% | 30% | 30% |
| Precision | 0% | 0% | 0% | 0% | 0% | 17% | 14% | 13% | 22% | 27% | 25% | 23% | 21% | 20% | |

| CONSULTA 2 | |
|--------------|-----|
| Media Google | 45% |
| Media Yahoo | 22% |

| | | | | | | | | | | |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| Yahoo | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| Google | 50% | 33% | 43% | 50% | 56% | 40% | 40% | 40% | 40% | 40% |
| Google | 17% | 22% | 27% | 27% | 27% | 27% | 27% | 27% | 27% | 27% |

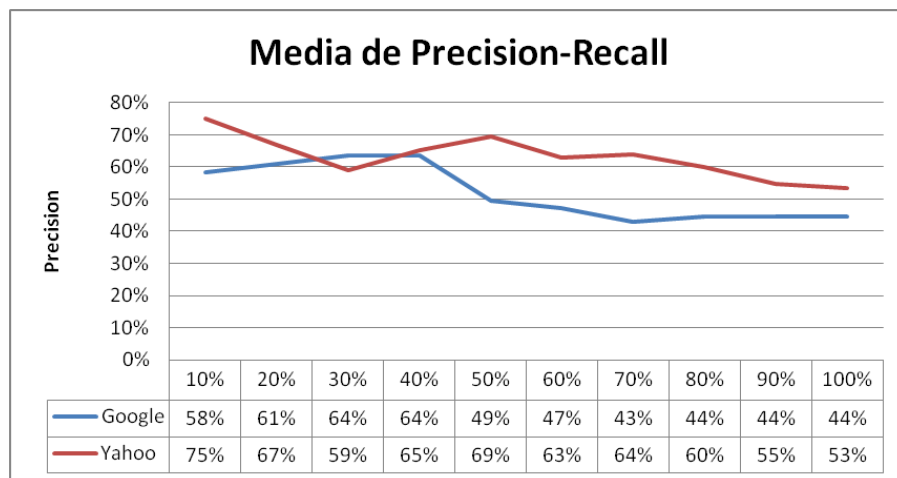




SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

Media

| | | | | | | | | | | | |
|------------|--------|------|------|------|------|-----|-----|-----|-----|-----|------|
| Consulta 1 | Google | 100% | 100% | 100% | 100% | 71% | 67% | 58% | 62% | 62% | 62% |
| Consulta 2 | Google | 17% | 22% | 27% | 27% | 27% | 27% | 27% | 27% | 27% | 27% |
| Consulta 1 | Yahoo | 100% | 100% | 75% | 80% | 83% | 86% | 88% | 80% | 69% | 67% |
| Consulta 2 | Yahoo | 50% | 33% | 43% | 50% | 56% | 40% | 40% | 40% | 40% | 40% |
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| Media | Google | 58% | 61% | 64% | 64% | 49% | 47% | 43% | 44% | 44% | 44% |
| Media | Yahoo | 75% | 67% | 59% | 65% | 69% | 63% | 64% | 60% | 55% | 53% |





SISTEMAS AVANZADOS DE RECUPERACIÓN DE INFORMACIÓN

Histograma

| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|------------|--------|------|------|------|------|-----|-----|-----|-----|-----|------|
| Consulta 1 | Google | 100% | 100% | 100% | 100% | 71% | 67% | 58% | 62% | 62% | 62% |
| Consulta 2 | Google | 17% | 22% | 27% | 27% | 27% | 27% | 27% | 27% | 27% | 27% |
| Consulta 1 | Yahoo | 100% | 100% | 75% | 80% | 83% | 86% | 88% | 80% | 69% | 67% |
| Consulta 2 | Yahoo | 50% | 33% | 43% | 50% | 56% | 40% | 40% | 40% | 40% | 40% |

R-Precision

| | Google | Yahoo | Comparación |
|------------|--------|-------|-------------|
| consulta 1 | 100% | 80% | 20% |
| consulta 2 | 27% | 50% | -23% |

