

# Síntesis de textos

Manuel Díaz Pastor

Mireia Pires State

Andrea Gascó Pau

Iñaki Diez Lambies

# Índice

## Introducción

- ¿En qué consiste la síntesis de textos?
- Objetivos.

## Evolución de las técnicas.

- Técnicas pioneras.
- Técnicas actuales.
- Técnicas utilizadas con sistemas inteligentes.

## Síntesis abstractiva de texto.

- Técnicas.
- Algunas implementaciones.

## Evaluación de las técnicas.

## Bibliografía

# Introducción

## Orígenes

- Luhn 1958
  - La creación automática de resúmenes de literatura

## Tecnologías del lenguaje humano

- HLT
- Comunicación máquina – humano
- Lenguaje natural

## Internet

- Mucha información
- Muchos formatos

¿En qué consiste la síntesis de textos?

# Objetivos

- Ahorrar tiempo a los usuarios.
  - Encontrar rápidamente información de forma concisa y relevante.
- Ahorrar tiempo a los sistemas de recuperación de la información.

# Retos

- Redundancia
- Dimensión temporal
- Correferencia
- Ordenación de frases
  - Multi-document summarization
- Coherencia
- Cambios en la sociedad
  - Nuevos requerimientos
  - Resúmenes basados en opiniones

# Evolución de las técnicas

Pioneras

Actuales

Con sistemas  
inteligentes

# Técnicas pioneras

## Estadística

Relevancia

Tema principal

Mayor puntuación

## Temas

Palabras clave

Firma de temas

## Grafos

Nodo por elemento

Conectividad

## Discurso

Estructura retórica

Relaciones sintácticas

## Aprendizaje automático

Clasificadores

Redes neuronales

Puntuación de oraciones



# Técnicas actuales

## Personalizados

Resúmenes basados en el perfil del usuario.  
Comparar preferencias del usuario con el texto.

## De actualización

Proporcionar información reciente, que el usuario desconoce.

## Basados en opiniones

Orientados a extraer los fragmentos que presentan la opinión transmitida en el texto.

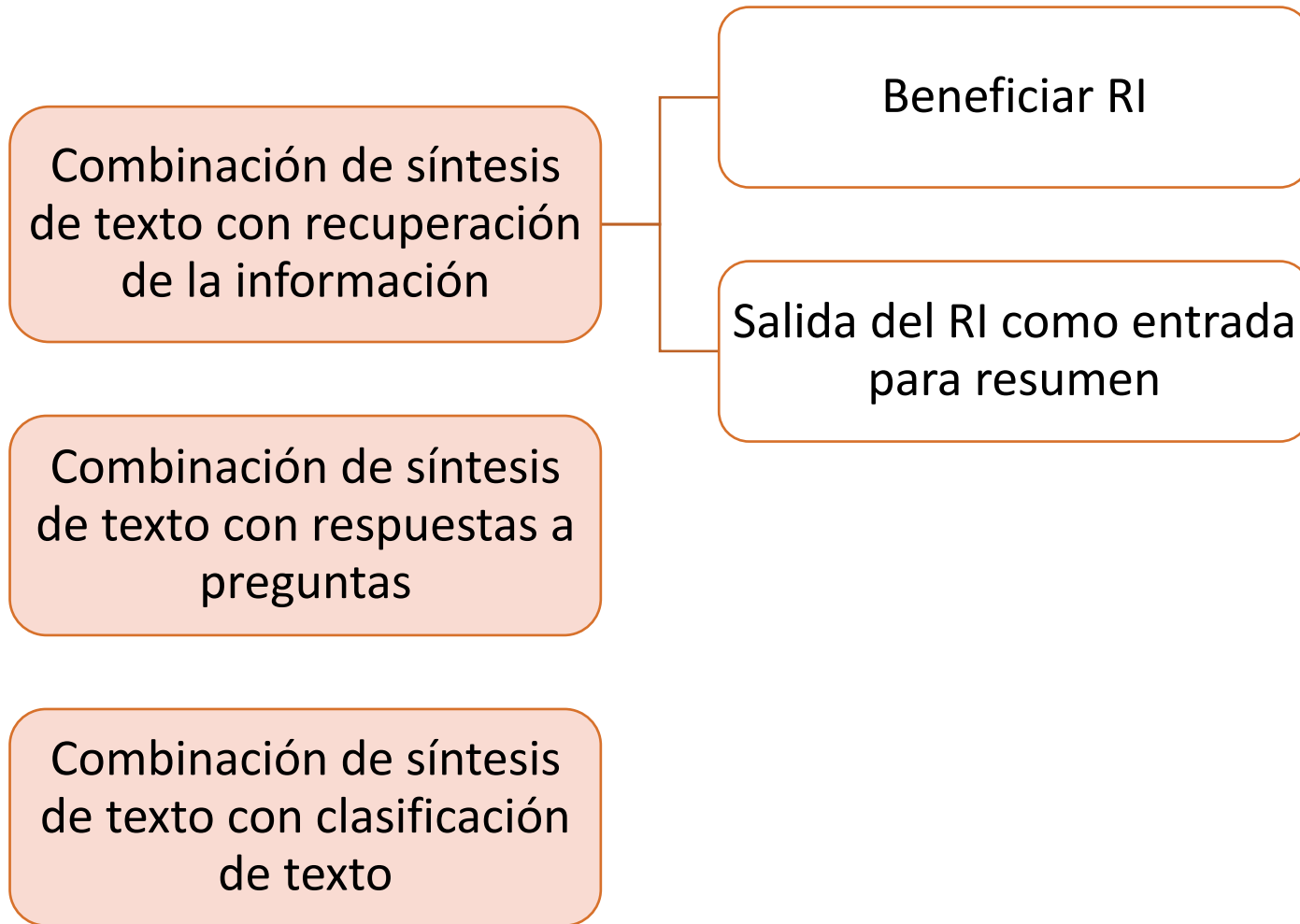
## De encuestas

Para extraer solo información esencial, por ejemplo, en biografías.

## Abstractos

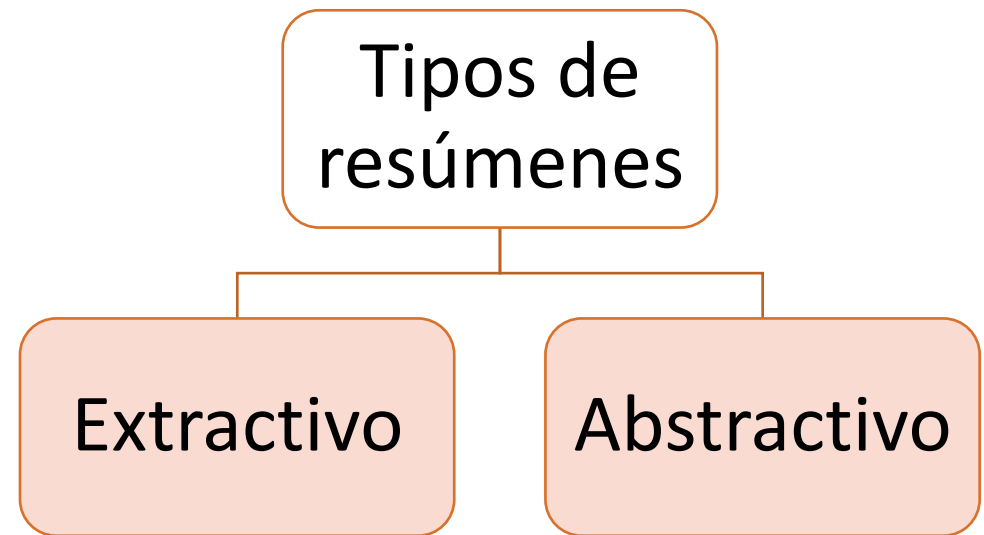
Generar textos con mayor coherencia mediante NLG, extrayendo y uniendo contenido.

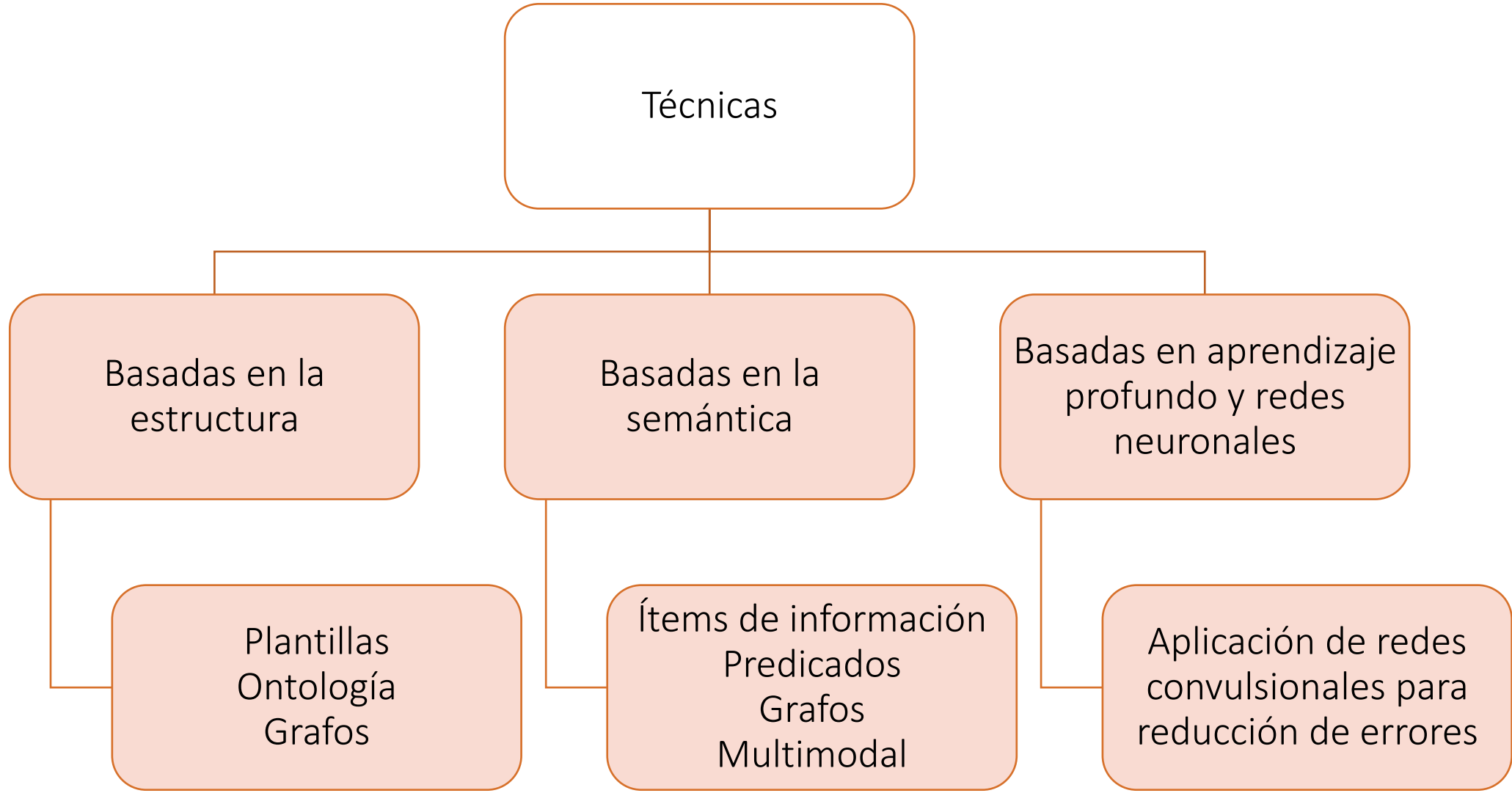
# Técnicas con sistemas inteligentes



**Objetivo de RI:** Recuperar documentos de naturaleza no estructurada (texto) que se encuentran en grandes colecciones de documentos y que cumplan una o varias condiciones.

# Síntesis abstractiva de texto





# Implementaciones

## Comprensión

- Reducir longitud
- Conservar significado

## Generalización

- Rutas de hiperonimia
- Información común

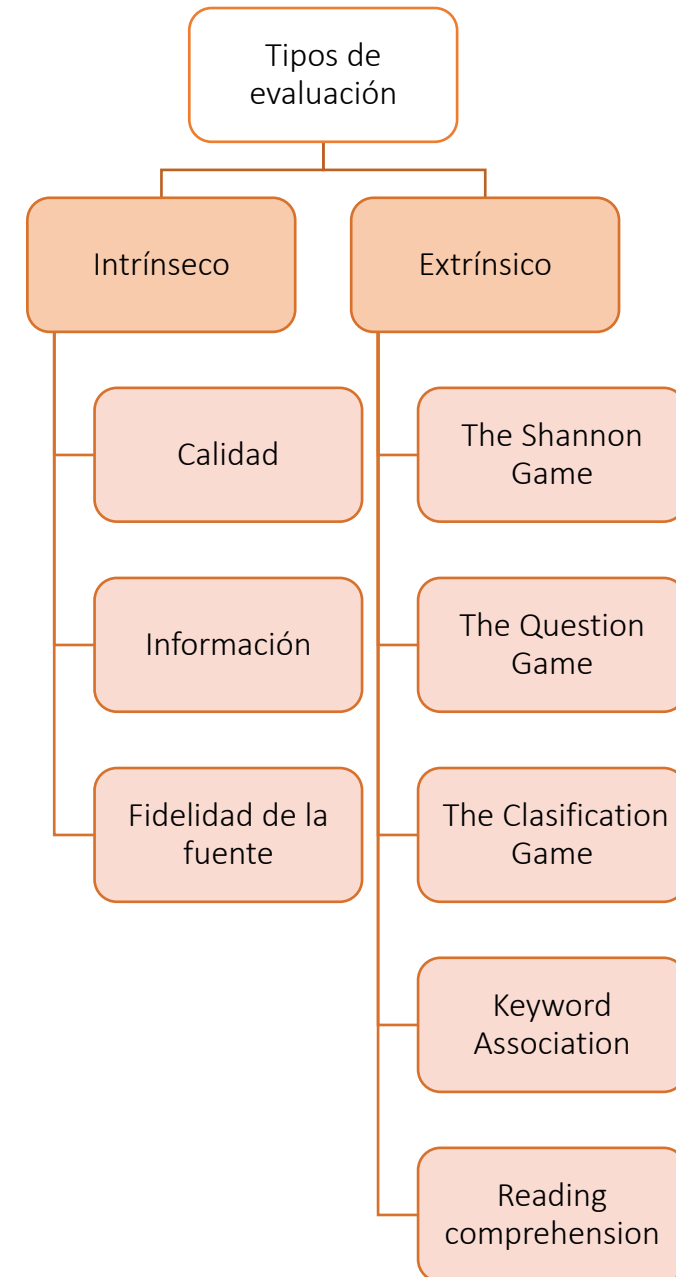
## Cálculo de puntuaciones

- Grafos semánticos
- Sujeto-objeto-verbo

## Generación

- Planificación
- Realización
- Evaluación

# Evaluación de las técnicas



# Evaluación de las técnicas

- Evaluación de la información

Métricas más conocidas: precisión, recall, F-media.

Al usar estas métricas se pueden juzgar mal dos resúmenes buenos para un mismo documento.

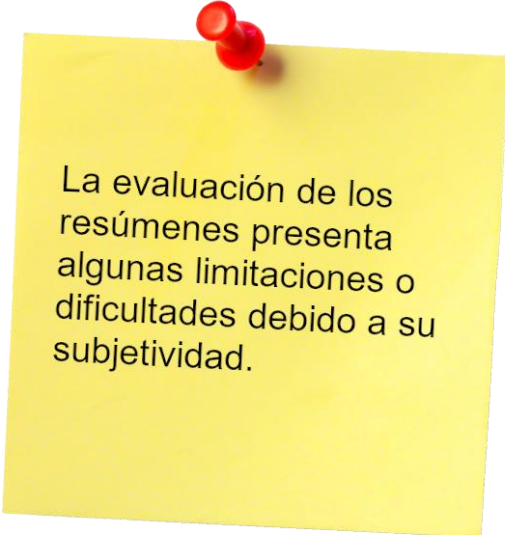
Se proponen:

- Utilidad relativa
- Factoid score
- Pyramid

- Evaluación de la calidad

Los métodos automáticos solo evalúan la calidad según su contenido, pero no por su coherencia o su redundancia.

- FAN Protocol
- MLUCE Protocol



La evaluación de los resúmenes presenta algunas limitaciones o dificultades debido a su subjetividad.

Conclusiones



# Bibliografía

---

Gupta, S., & Gupta, S. K. (2019). Abstractive summarization: An overview of the state of the art. *Expert Systems with Applications*, 121, 49-65.

---

Lloret, E., & Palomar, M. (2012). Text summarisation in progress: a literature review. *Artificial Intelligence Review*, 37(1), 1-41.

---

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2), 159-165.

---

Sakai, T., & Sparck-Jones, K. (2001, September). Generic summaries for indexing in information retrieval. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 190-198).

---

Napoles, C., Callison-Burch, C., Ganitkevitch, J., & Van Durme, B. (2011, June). Paraphrastic sentence compression with a character-based metric: Tightening without deletion. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation* (pp. 84-90).