



Universidad Autónoma del Estado de México



Centro Universitario UAEM Valle de México

Ingeniería en Sistemas y Comunicaciones

Unidad de Aprendizaje

Temas Selectos de Sistemas

Unidad de Competencia

Aplicaciones

Elaboró: Saturnino Job Morales Escobar

Fecha de elaboración: Septiembre de 2018



Programa de Estudios por Competencias
TEMAS SELECTOS DE SISTEMAS

I. IDENTIFICACIÓN DEL CURSO

ORGANISMO ACADÉMICO: CENTRO UNIVERSITARIO UAEM VALLE DE MEXICO								
Programa Educativo: INGENIERIA EN SISTEMAS Y COMUNICACIONES					Área de docencia: CIENCIAS DE LA INGENIERÍA			
Aprobación por los H. H. Consejos Académico y de Gobierno			Fecha: Fecha de actualización:		Programa elaborado por: SATURNINO JOB MORALES ESCOBAR, IVONNE RODRÍGUEZ PÉREZ		Fecha de actualización: Noviembre 2010	
Clave	Horas de teoría	Horas de práctica	Total de horas	Créditos	Tipo de Unidad de Aprendizaje	Carácter de la Unidad de Aprendizaje	Núcleo de formación	Modalidad
L32295	3	1	4	7	CURSO	OBLIGATORIA	SUSTANTIVO	PRESENCIAL
Prerrequisitos (Conocimientos Previos): PROBABILIDAD Y ESTADÍSTICA PARA INGENIEROS, MATEMÁTICAS DISCRETAS, ALGORITMOS Y ESTRUCTURAS DE DATOS.					Unidad de Aprendizaje Antecedente NINGUNA		Unidades de Aprendizaje Consecuentes: NINGUNA	
Programas educativos en los que se imparte: INGENIERIA EN SISTEMAS Y COMUNICACIONES								

CIENCIAS BÁSICAS Y MATEMÁTICAS

ALGEBRA Y GEOMETRÍA ANALÍTICA	3 3 9	ALGEBRA LINEAL	2 2 6	ECUACIONES DIFERENCIALES	3 1 7	CÁLCULO VECTORIAL	3 1 7
ESTÁTICA Y DINÁMICA	3 3 9	CÁLCULO DIFERENCIAL E INTEGRAL	3 1 7	LÓGICA MATEMÁTICA	3 1 7	ELECTROMAGNETISMO	2 2 6
QUÍMICA	2 2 6	MATEMÁTICAS DISCRETAS	3 1 7	MÉTODOS NUMÉRICOS	4 2 10	PROBABILIDAD Y ESTADÍSTICA PARA INGENIEROS	4 2 10

CIENCIAS DE LA INGENIERÍA

INTRODUCCIÓN A LA INGENIERÍA	2 4 8	FUNDAMENTOS DE PROGRAMACIÓN	3 3 9	FUNDAMENTOS DE BASES DE DATOS	3 1 7	SISTEMAS OPERATIVOS	4 2 10	LENGUAJES FORMALES Y AUTÓMATAS	2 2 6	ELECTRÓNICA ANALÓGICA	4 2 10	COMUNICACIÓN POR MEDIOS ÓPTICOS 226	TEMAS SELECTOS DE SISTEMAS	3 1 7
LENGUAJES DE BAJO NIVEL	2 4 8	ALGORITMOS Y ESTRUCTURAS DE DATOS	2 4 8	BASES DE DATOS	2 4 8	PROGRAMACIÓN ORIENTADA A OBJETOS	2 4 8	INGENIERÍA DE SOFTWARE	2 4 8	COMUNICACIÓN VÍA MICROONDAS Y SATELITAL 226	SISTEMAS ELECTRÓNICOS DE COMUNICACIÓN 226	COMPLADORES 226		
CIRCUITOS ELÉCTRICOS	2 2 6	SISTEMAS DE INFORMACIÓN	2 4 8	PROGRAMACIÓN AVANZADA 226		TEORÍA DEL CONTROL* 226		GRAFICACIÓN 226						
INVESTIGACIÓN DE OPERACIONES	2 4 8	ADMINISTRACIÓN DE BASES DE DATOS 226		CAIDAD DEL SOFTWARE 226										
SISTEMAS OPERATIVOS PARA RED	4 2 10													

INGENIERÍA APLICADA

INTRODUCCIÓN A LA COMPUTACIÓN	2 4 8	ARQUITECTURA DE COMPUTADORAS	2 4 8	REDES	4 2 10	PROTOSCOLOS DE COMUNICACIÓN DE DATOS	3 3 9	DESARROLLO DE PROYECTOS	2 2 6	FORMULACIÓN Y EVALUACIÓN DE PROYECTOS	3 3 9	RESIDENCIA PROFESIONAL	0 30 30
SEMINARIO DE TITULACIÓN* 226		INTELIGENCIA ARTIFICIAL* 226		SISTEMAS DIGITALES	2 4 8	SISTEMAS DISTRIBUIDOS 226		INTERCONEXIÓN Y SEGURIDAD EN REDES	2 2 6	TRANSMISIÓN Y COMUNICACIÓN DE DATOS 226		SISTEMAS EXPERTOS 226	SISTEMAS DE INSTRUMENTACIÓN Y CONTROL 226
SISTEMAS DE TIEMPO REAL 226								TALLER DE INVESTIGACIÓN 226					

CIENCIAS SOCIALES Y HUMANIDADES

ADMINISTRACIÓN 317	CONTABILIDAD 317	ECOLOGÍA, ÉTICA Y NORMATIVIDAD 317	INGLÉS C1	INGLÉS C2	PSICOLOGÍA ORGANIZACIONAL 226	PLANEACIÓN ESTRATÉGICA 226	METODOLOGÍA DE LA INVESTIGACIÓN	2 2 6
CIENCIA Y HUMANISMO 317	PROBLEMAS SOCIOECONÓMICOS DE MÉXICO 317	TÉCNICAS DE COMUNICACIÓN 317			ADMINISTRACIÓN DE CENTROS DE CÓMPUTO	2 2 6	AUDITORIA Y SEGURIDAD INFORMÁTICA 226	

Temas Selectos de Sistemas

Unidades de Competencia

- 1. Conceptos Básicos del Reconocimiento de Patrones*
- 2. Selección de Variables*
- 3. Clasificación Supervisada*
- 4. Clasificación no Supervisada*
- 5. Procesamiento de Lenguaje Natural*
- 6. Aplicaciones*

Evaluación de la unidad de competencia

<i>Exámenes</i>	<i>50%</i>
<i>Exposición de una aplicación</i>	<i>10%</i>
<i>Ejercicios</i>	<i>40%</i>

Objetivo de la unidad de competencia

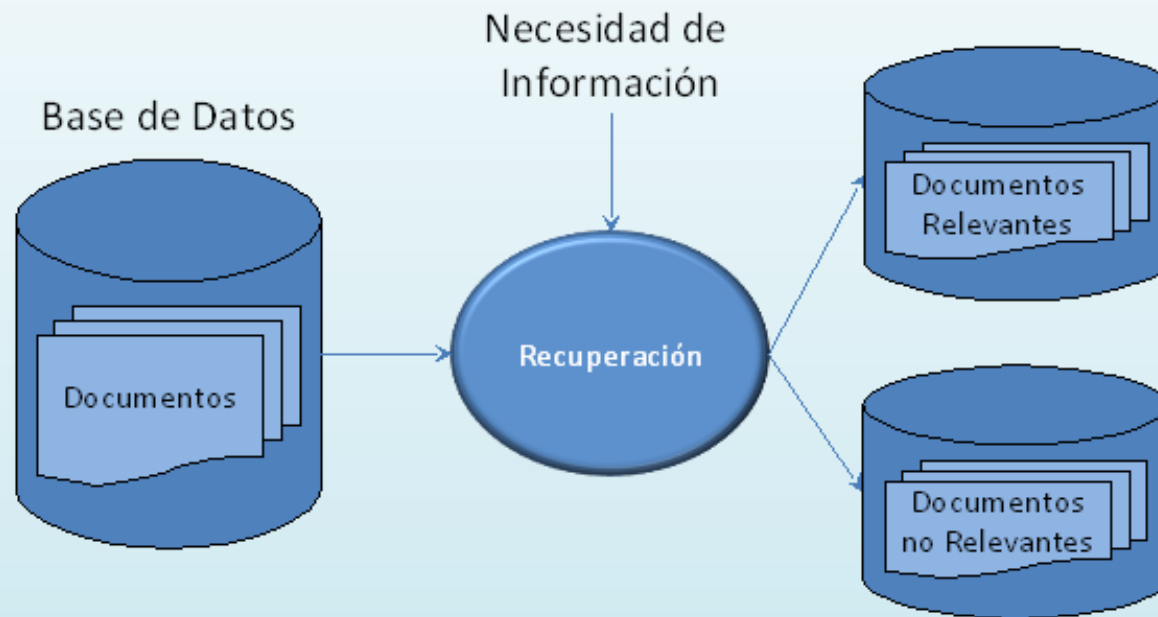
Analizar la aplicación de los algoritmos y técnicas del Reconocimiento de Patrones y el procesamiento del lenguaje natural a la minería de datos y a los sistemas de recuperación de información.

En la presentación anterior se abordó la aplicación a la minería de datos



En esta presentación se abordarán los sistemas de recuperación de información

Sistemas de Recuperación de Información



(19)

Contenido

- *Introducción*
- *Conceptos sobre Sistemas de Recuperación de Información*
- *Presentar un panorama de los sistemas de recuperación de información*
- *Métodos utilizados en los sistemas de recuperación de información*
- *Aplicación*

Introducción

Ante la gran cantidad de datos, documentos, información, imágenes, etcétera, que se generan de manera digital, la capacidad de procesarla manualmente es sobrepasada y su acumulación da como resultado un problema humanamente intratable.

Una solución, es desarrollar herramientas computacionales que realicen ese trabajo.



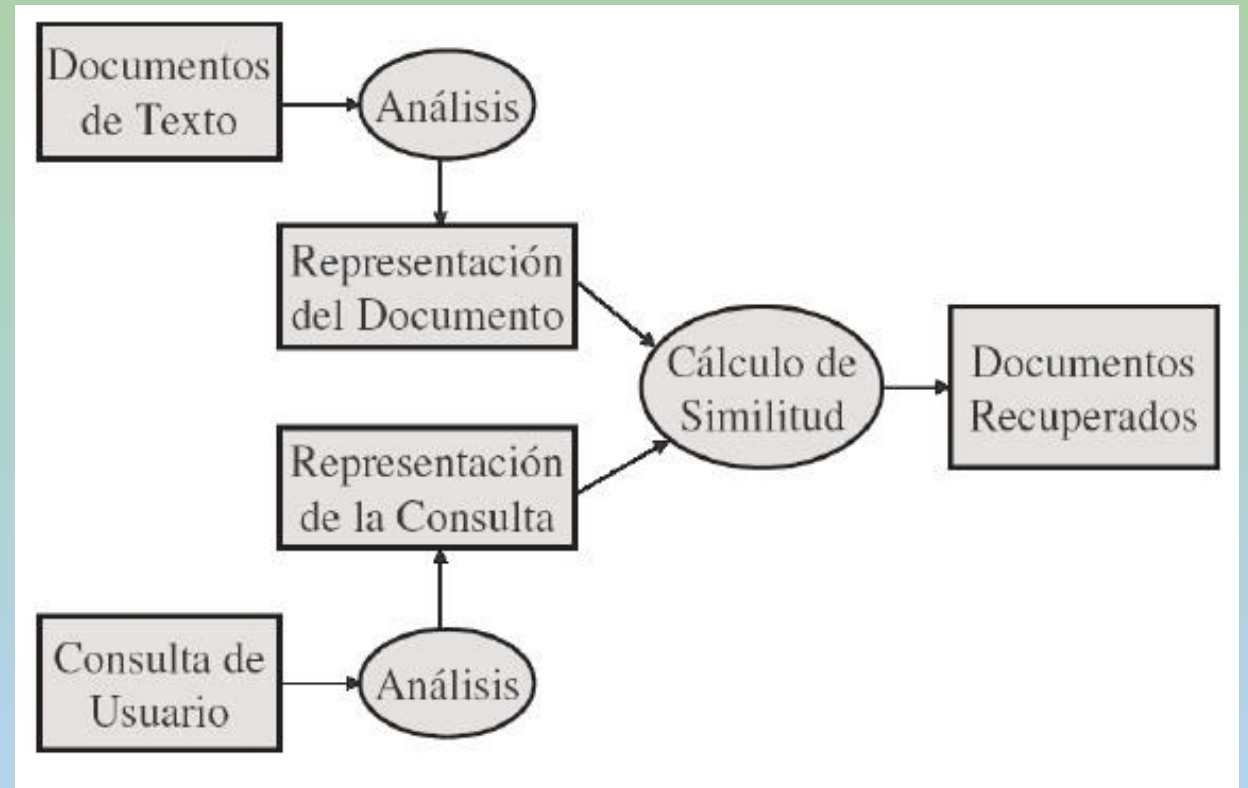
Bases de datos contra un Sistema de Recuperación de Información

SGBD	SRI
<ul style="list-style-type: none">• Información estructurada• Recuperación determinista• Consulta específica• Lenguaje formal y estructurado• Aciertos exactos	<ul style="list-style-type: none">• Información semi o no estructurada• Recuperación probabilística• Consulta imprecisa• Lenguaje natural y no estructurado• Aciertos parciales

Introducción

continuación

Así, se han desarrollado diversos tipos de sistemas, desde sistemas de información hasta sistemas que permitan obtener datos no estructurados contenidos en documentos de texto.



<http://recupinfo2011.blogspot.mx/2011/03/sistemas-de-recuperacion-de-informacion.html>

Tareas Básicas de un Sistema de Recuperación de Información

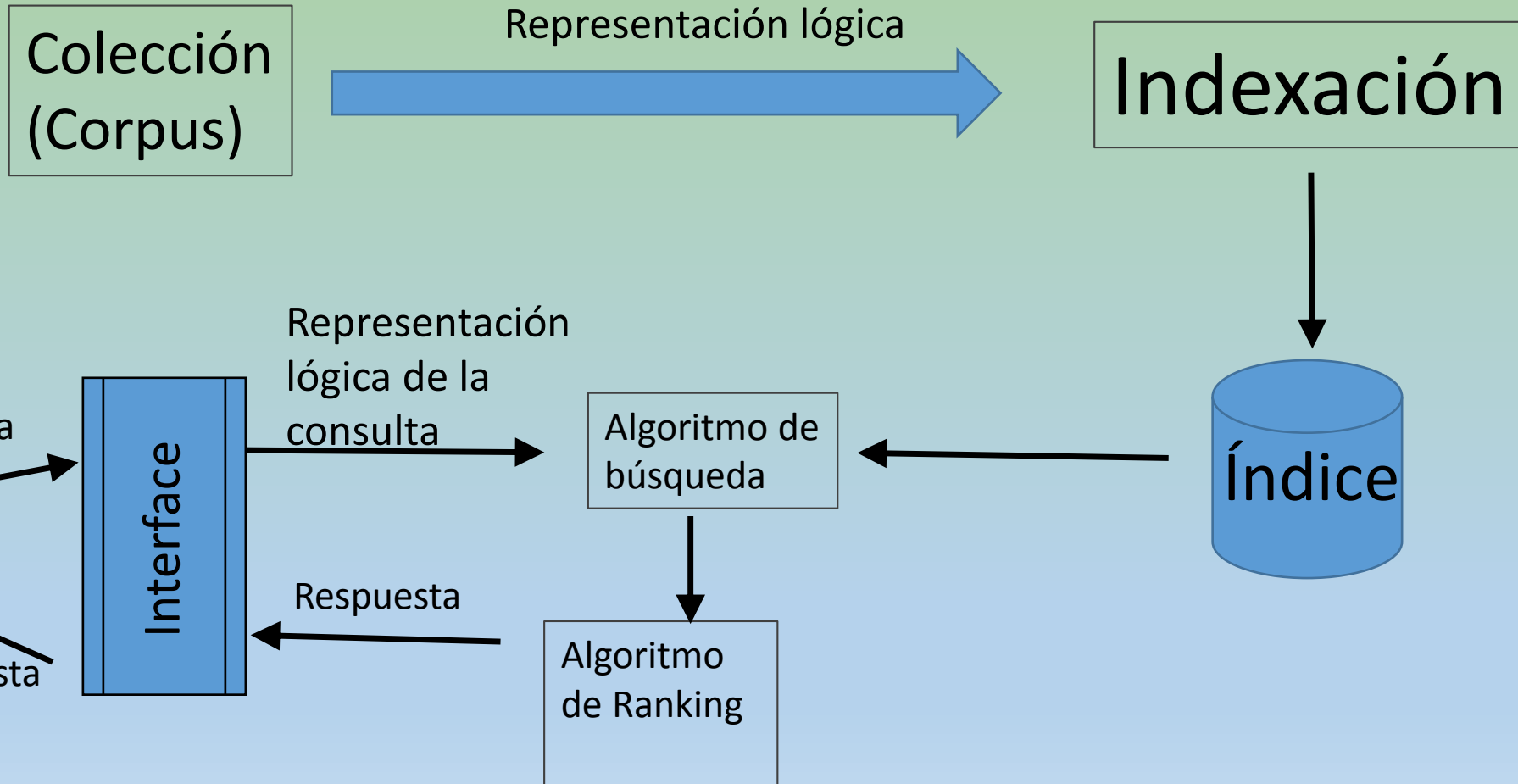
- *Representación:*

La representación de información es fundamental, de ello depende lo que se pueda recuperar mediante la búsqueda.

- *La necesidad recuperar información es equivalente a realizar una consulta.*

En los SRI, la recuperación de información consiste en determinar cuáles documentos de una colección contienen las palabras claves de una pregunta del usuario que cubren sus necesidades de información.

Arquitectura de un SRI



Descripción de los componentes de un SRI

- *Colección: Documentos de textos (sucesión de cadenas, palabras) escritos en lenguaje natural (corpus o base de datos textual).*
- *Representación lógica: Representación de todos los documentos para poder realizar operaciones, puede consistir de términos, frases u otras unidades (sintácticas o semánticas) que los caracterizan.*
- *La representación de documentos mediante un conjunto de términos se le conoce como “bolsa de palabras”*

Descripción de los componentes de un SRI

- *Considerando la representación lógica de los documentos se realiza un proceso de indexación que permitirá la construcción de las estructuras de datos (denominada índices) que la almacena.*
- *Las estructuras permitirán realizar búsquedas eficientes.*
- *El algoritmo de búsqueda acepta como entrada una expresión de consulta de un usuario y verificará en los índices cuáles documentos pueden satisfacerlo.*

Descripción de los componentes de un SRI

- *EL algoritmo de ranking determinará la relevancia de cada documento y retornará una lista de posibles respuestas.*
- *El primer ítem de la lista corresponde al documento más relevante y de manera sucesiva serán considerados en orden decreciente*
- *La interface del usuario permite que especifique su consulta mediante una expresión escrita en un lenguaje definido y sirve para mostrar las respuestas que ofrece el sistema.*

Tipos de recuperación de información

- *Recuperación inmediata*

El usuario planea la necesidad de información y recibe las referencias a los documentos.

- *Búsqueda propiamente dicha o recuperación ad-hoc*

El usuario formula su consulta en lenguaje natural y el sistema la evalúa y responde. Un ejemplo común son los buscadores de internet como Google, yahoo, etc.

- *Navegación o browsing*

El sistema ofrece una interface con temas donde el usuario navega y obtiene referencias a los documentos relacionados.

Tipos de recuperación de información

- *Recuperación diferida*

El usuario especifica sus necesidades y el sistema entregará la forma continua los nuevos documentos que lleguen y cumpla con ésta.

- *Filtrado y ruteo*

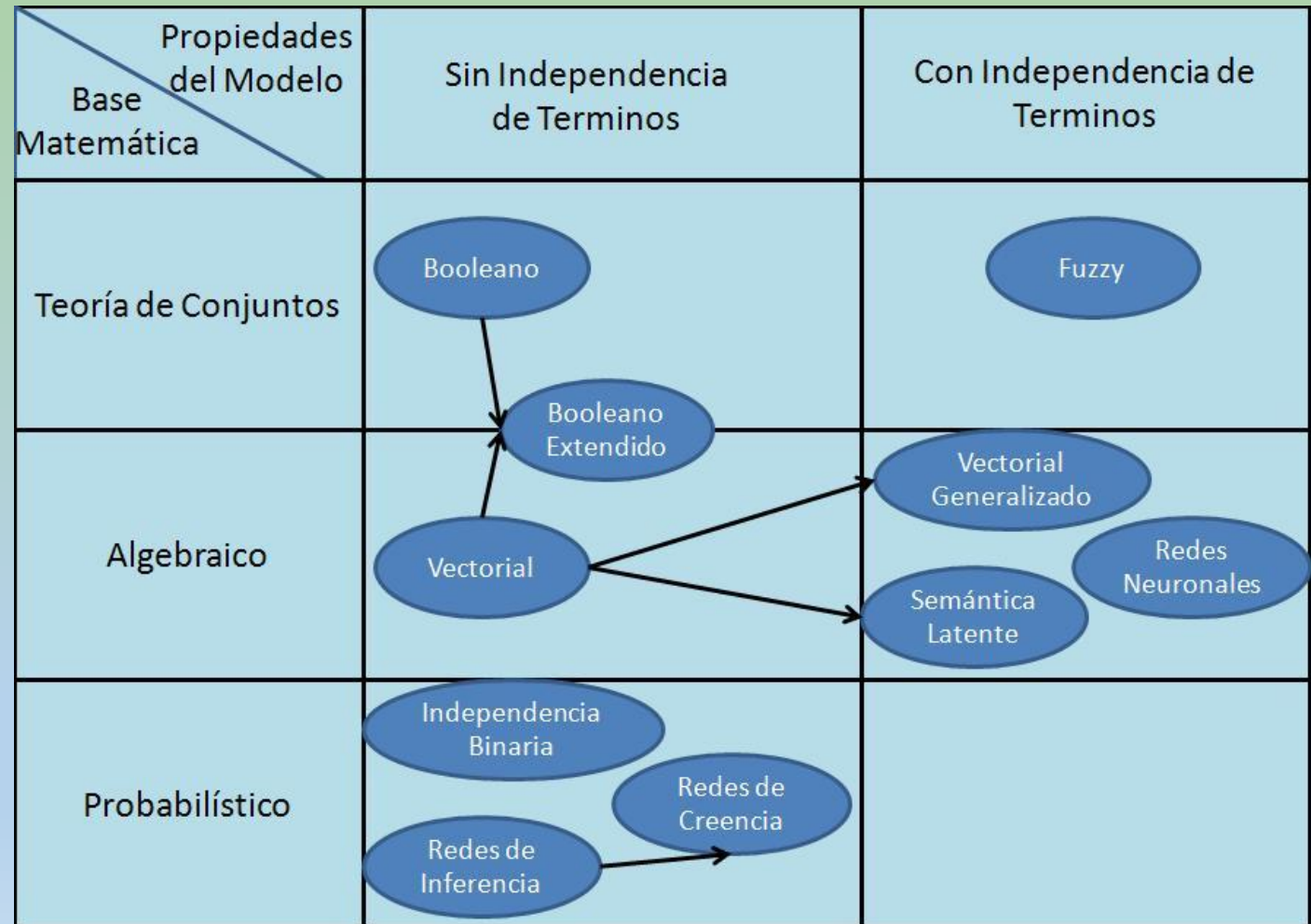
La necesidad del usuario define un “perfil” (profile) de los documentos buscados.

Cada vez que un documento llega el sistema lo compara con el perfil y de ser relevante lo envía al usuario.

La consulta es relativamente estática y el usuario tiene un rol pasivo. La dinámica la da la aparición de nuevos documentos.

Recuperación de Información

La recuperación de información es un estudio interdisciplinario.



Actividades para gestión y organización de conocimiento explícito

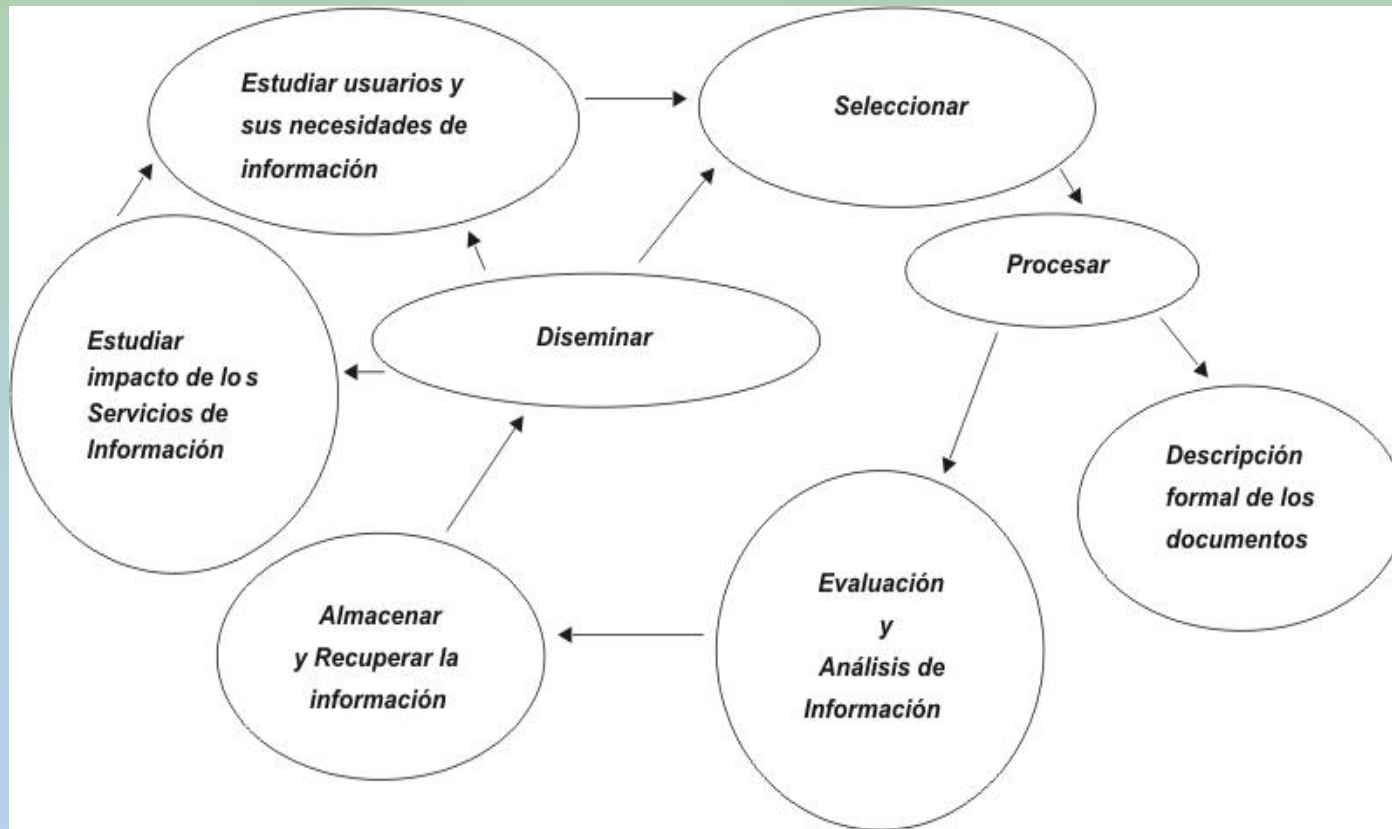


Fig. Tareas básicas del profesional de la información para la gestión y organización del conocimiento explícito.

Actividades para un sistema tradicional de recuperación de información.

<http://scielo.sld.cu/img/revistas/aci/v12n2/f0112204.jpg>

Modelos de Recuperación de Información

- *Recuperación*

- *Modelos clásicos*

- *Booleano*
 - *Vectorial*
 - *Probabilístico*

- Modelos Estructurados*

- *Listas no solapadas*
 - *Nodos proximales*

Modelos de Recuperación de Información cont.

- *Navegación*

- *Plana*

- *Guiada por estructura*

- *Hipertexto*

Procesamiento de Lenguaje natural

Es un campo interdisciplinario

- *Ciencias Computacionales*
- *Inteligencia Artificial*
- *Lingüística Computacional*



Propósito

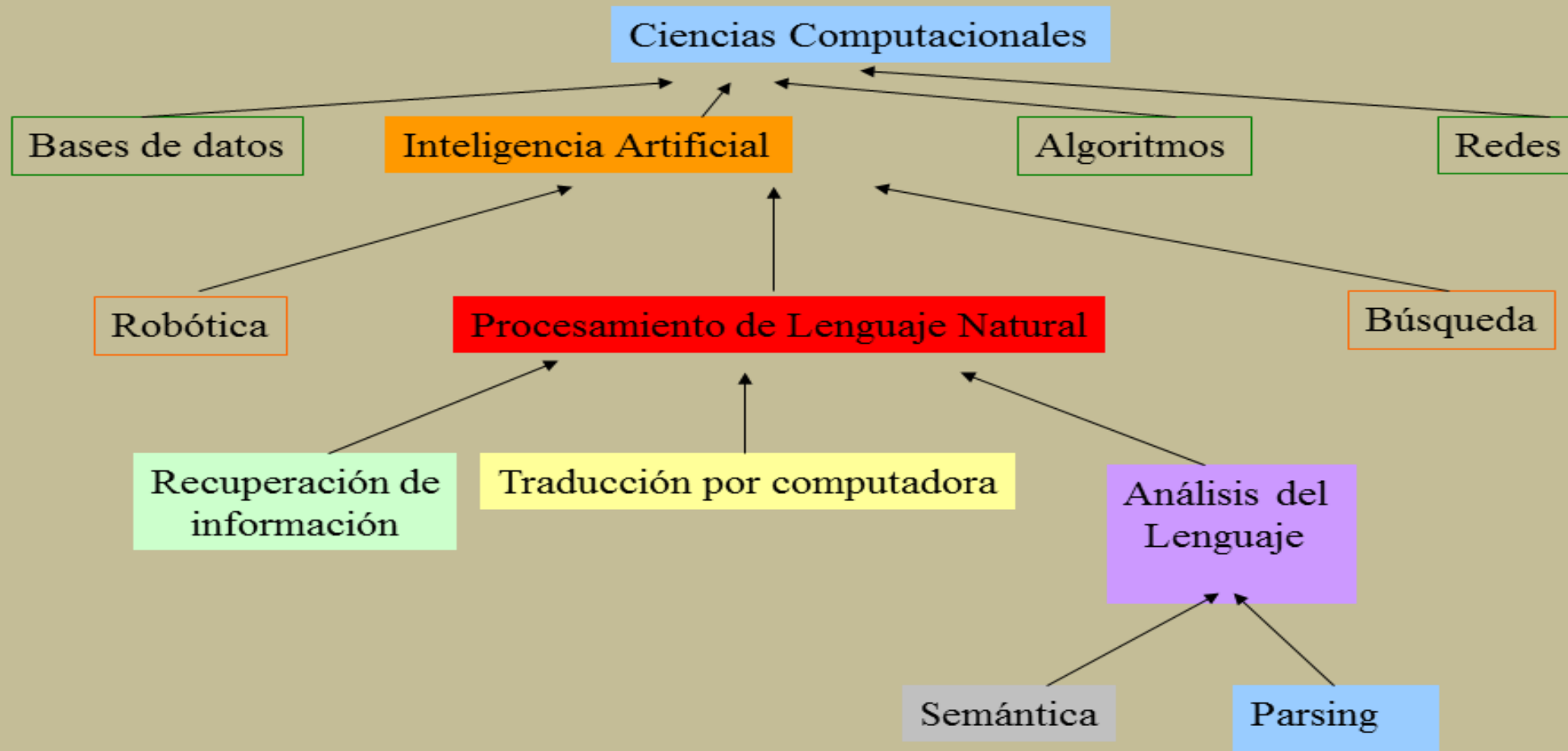
Que la computadora entienda el lenguaje natural, hablado o escrito

https://www.google.com.mx/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&ved=2ahUKewj7qbHgh9_dAhVR4qwKHbM4BmwQjRx6BAgBEAU&url=http%3A%2F%2Fcomunicacionycomputadoras.blogspot.com%2F2016%2F03%2Fla-computadora-como-medio-de.html&psig=AOvVaw3JLYgEg1MgFgN2yHwl9wpO&ust=1538270841525653

Importancia de aplicar el procesamiento de lenguaje natural

- *El repositorio de conocimiento humano está conformado por textos y crece de manera exponencial.*
- *Se requieren herramientas para analizar, entender, generar lenguaje humano como lo realizan los humanos*
- *En la siguiente imagen, se ubica esta área con respecto a las ciencias computacionales y su relación con la recuperación de información (<https://www.monografias.com/docs113/analisis-textos-analisis-semantico/Diapositiva4.png>)*

¿Cómo se ubica en la taxonomía de las CC?



APLICACIONES DE LAS TECNOLOGÍAS DEL LENGUAJE

Enseñanza/Aprendizaje del Lenguaje



Recuperación de Información & Extracción de Información

Sistemas para entender historias

Sumarización (resúmenes) Automática

Categorización/Ruteado de Texto

Detección de Plagio

Clasificación de Documentos

Texto-a-voz & Reconocimiento de voz

Traducción automática (asistida)

Análisis de documentos en imágenes

Question answering
Respuestas automáticas

Interfaces de Diálogo de Lenguaje Natural a Bases de Datos

Corrección de Textos—Ortografía & Gramática

Aplicación de las tecnologías del lenguaje a PLN

<https://www.monografias.com/docs113/analisis-textos-analisis-semantico/Diapositiva17.png>

Aplicación de las tecnologías del lenguaje a PLN

Escalabilidad de los lenguajes

Gran cantidad de textos en la WEB (Big Data)

- *¿Cómo manejarla?*
- *¿Qué clase de información se puede extraer?*

Aplicación de las tecnologías del lenguaje a PLN

Enfoques para extraer información

- *Técnicas secuenciales o paralelas*
- *Métodos de aprendizaje supervisado*
- *Técnicas de minería de datos*

Análisis semántico e interpretación

Procesar lenguaje para producir conocimiento de sentido común

Obtener información valiosa de metadatos

LA Representación de conocimiento

Identificación de información relevante

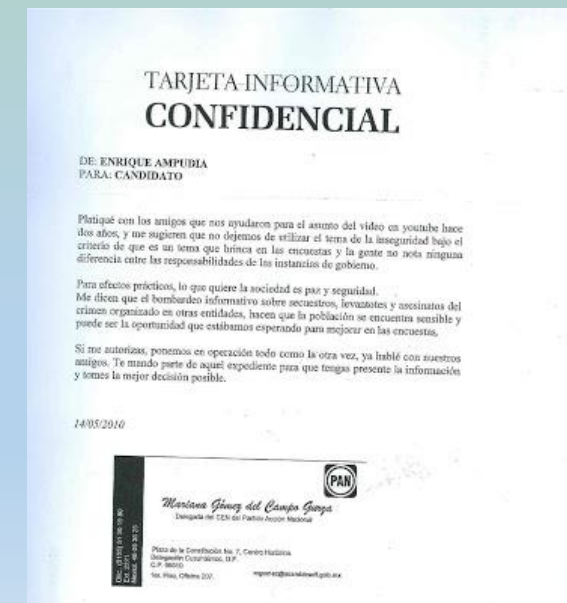
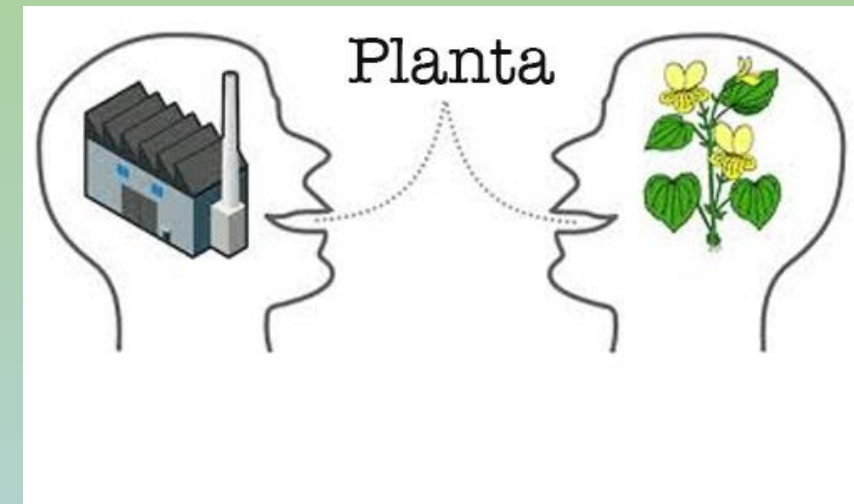
Técnicas para la representación de información

Modelos de grafos

Modelos de espacios vectoriales

Ejemplos de aplicación de análisis semántico

- *Identificación de autor*
- *Minería de opiniones*
- *Similitud de textos y respuestas a preguntas*
- *Detección de unidades semánticas mínimas y su significado*
- *Determinación de sensibilidad de documentos*



Se recomienda visitar las siguientes ligas para profundizar en el tema de enfoques para atacar problemas de Procesamiento de Lenguaje Natural

Principales competencias internacionales de investigación para evaluar nuevos enfoques para atacar problemas de Procesamiento de Lenguaje Natural:

❖ **CLEF Pan (Europa)*:**

- Una serie de eventos científicos y tareas compartidas sobre el análisis de texto digital y forense.
- Diferentes problemas como: Detección de plagio, Identificación de autor, Identificación de Acoso Sexual, etc.

*<http://pan.webis.de/>



❖ **SemEval (Norteamérica)**:**

- Una serie de evaluaciones de análisis de sistemas de semántica computacional.
- Diferentes problemas como: desambiguación del sentido de las palabras, Pregunta-respuesta (Question Answering), Análisis de sentimientos, etc.

** <http://alt.qcri.org/semEval2016/>



Aplicación para verificación de autor

Uso de grafo de flujo sintáctico

Vértices

Todas las palabras en un corpus de entrenamiento excepto las palabras vacías (llamadas stop words)

Ligas

Dos vértices están conectados por medio de la secuencia del texto

Dos vértices están conectados si están al principio y al final de la frase

Se considera la estructura y secuencia sintáctica de oraciones

Aplicación para verificación de autor

Uso de medidas de centralidad para extraer frases relevantes:

Chuncks

Cliques

Uso de un método de aprendizaje supervisado

Uso de corpus especiales y conjuntos de datos preparados para esa tarea

Bibliografía

- *Soumen Chakrabarti.*

“Mining the Web : discovering knowledge from hypertext data / Soumen Chakrabarti.”

Morgan Kaufmann Publisher, 2003

- *Mehmed Kantar*

“Data Mining Concepts Models Methods and Algorithms”

Hoboken, N.J. : John Wiley : IEEE Press, 2011

- *Anne Kao and Stephen R. Poteet (eds)*

“Natural language processing and text mining”.

London : Springer, 2010.

Bibliografía

- *Ian H. Witten, Eibe Frank, Mark A. Hall.*
“Data mining: practical machine learning tools and techniques”, 3rd ed.
Amsterdam ; Boston, MA : Morgan Kaufman : Elsevier, c2011.
- *Nakov, p. & Zesch, T. Computational semantic analysis of language (2016)*

Guión Explicativo

- *Este material está desarrollado como apoyo al curso presencial de la unidad de competencia “Aplicaciones” de la UDA “Temas Selectos de Sistemas”, correspondiente a la carrera de Ingeniería en Sistemas y Comunicaciones.*
- *Esta unidad de competencia es la continuación de la unidad de competencia Procesamiento de Lenguaje Natural.*

Guión Explicativo

- *Se recomienda estudiar el tema antes de la sesión presencial.*
- *Se recomienda seguir la secuencia en la que se presenta el material para mayor comprensión de los temas.*
- *Los temas se pueden consultar en extenso en las referencias proporcionadas en la bibliografía.*

Guión Explicativo

- *Se recomienda realizar ejercicios adicionales para complementar lo visto en clase y el material presentado.*
- *Al final de la unidad, se expondrá un artículo científico de reciente publicación donde se ilustre el desarrollo de alguna técnica o su aplicación.*