

El Resumidor

Una herramienta informática para extraer conocimiento desde textos en lenguaje natural.

Esta presentación está basada en los siguientes trabajos:

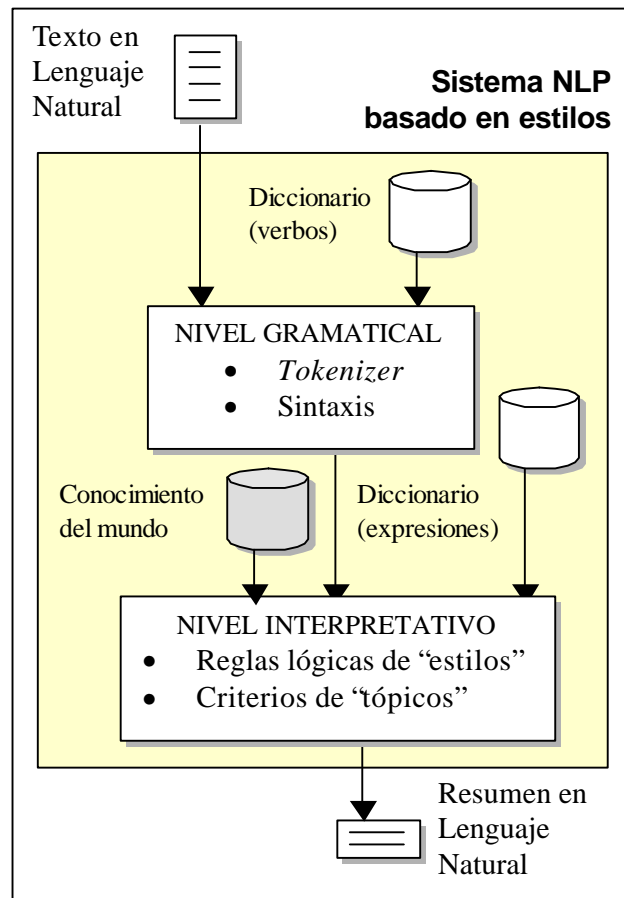
Ochoa-Arias, Jesús I. **Formulación de Normas Lógicas para el Discurso Escrito de Noticias.** Tesis. Escuela de Ingeniería de Sistemas. Universidad de Los Andes. 1999.

Contreras, H. Yelitza. **Una Técnica para la Extracción Automática de Resúmenes basada en una Gramática de Estilo.** Tesis. MSc Computación. Universidad de Los Andes. 2002.

Parra, M. Marilú. **Un Modelo Computacional para la Generación de Resúmenes Automáticos de Artículos Científicos en Español.** Tesis. MSc Modelado y Simulación. Universidad de Los Andes. 2004.

Arteaga-Matute, Leonardo. **Estrategia para el Procesamiento del Lenguaje Natural Español escrito en páginas Web con un Resumidor de Textos Simbólicos basado en Estilos.** Tesis. Licenciatura en Computación. La Universidad del Zulia. 2008.

El resumidor de textos basado en estilo (Contreras)



El resumidor de textos basado en estilo

Para ilustrar la estrategia, se utiliza como discurso el texto de una noticia internacional “The Wall Street Journal Americas” (15-02-2001). En este ejemplo se aplica “tópico común más específico”^[1] en todo el texto (con tópico entendido Williams).

El texto del discurso es el siguiente: “*Un informe de un comité científico de la unión que las ovejas y las cabras pueden contraer, teóricamente, el mal de las vacas hasta ahora esto sólo ha ocurrido en experimentos de laboratorio*”.

Las reglas de extracción son las siguientes:

- R1- T es un tópico del discurso D **si** en el discurso D, un Agente *revela* T
- R2- T es un tópico del discurso D **si** en el discurso D, un Agente *revela* T’ información que Agente2 *puede* hacer T.
- R3- T es un tópico del discurso D **si** en el discurso D, un Agente *revela* T’ información que Agente2 *puede contraer* Algo y T = “Algo en Agente”.
- R4- Un tópico T es el común más específico en D **si** es un tópico del discurso D tópico T’ de D tal que T’ mas general que T.
- R5- T’ es más general que T **si** T’ es más breve que T.

Con estas reglas el tópico T resultante es: “*mal de las vacas locas en ovejas y cabras*”

^[1] Denominamos “tópico común más específico” a aquel tópico que es menos general. La generalidad la determinamos en función de la longitud que constituye el tópico, por lo tanto un tópico más general es más breve

El resumidor

<http://resumidor.sourceforge.net>

El resumidor

- Objetivo de la investigación :

“El propósito general de este proyecto es crear un modelo de la estructura de un texto que podamos usar luego para guiar un sistema de cómputo que aproveche esa estructura para analizar y sintetizar (resumir) un texto” [Parra, 2004].

Lingüística del texto

- **Microestructura** [Williams, 1990] [Contreras, 2002]
 - Nivel-1
 - Nivel-2
- **Macroestructura**
 - Omisión y selección [van Dijk, 1983]
- **Superestructura** [Parra, 2004]
 - Esquema del texto

Lingüística del Texto

- **Modelo IMRD:**
 - La sección introducción de los AIC
 - Modelo *Create a Research Space-CARS*
[Swales, 1990]
 - La sección discusión de los AIC

Modelo CARS

Estructura retórica – Modelo Swales	
Movimiento1 Estableciendo el territorio	<i>Paso1</i> Centralización del tema <i>Paso2</i> Generalización del tópico <i>Paso3</i> Referencias a investigaciones previas
Movimiento2 Estableciendo el nicho	<i>Paso 1A</i> Contradicción con la tradición o <i>Paso 1B</i> Marcación de una brecha o <i>Paso 1C</i> Formulación de preguntas o <i>Paso 1D</i> Continuación con la tradición
Movimiento3 Ocupando el nicho	<i>Paso 1A</i> Presentación de objetivos o <i>Paso 1B</i> Anuncio de la investigación <i>Paso 2</i> Anuncio de los hallazgos <i>Paso 3</i> Presentación de la estructura

Movimiento3 paso1B

	<i>“Este trabajo tiene el propósito...”</i>
Movimiento3	<i>“La finalidad de nuestro trabajo es analizar...”</i>
Ocupando el nicho	<i>“El presente trabajo indica...”</i>
PASO 1B	<i>“nuestro objetivo es identificar...”</i>
	<i>“El presente artículo tiene por objetivo el desarrollo...”</i>

DCG para Movimiento3

DCG: *sentencia(sent(FN,FV))--> frase_nominal(FN),frase_verbal(FV).*

```
sentencia --> movimiento3.
```

```
movimiento3 --> frase_de_proposito, pronombre, copula,
```

```
verbo_proposito.
```

```
frase_de_proposito --> articulo, proposito, de, trabajo.
```

```
articulo --> [el].
```

```
proposito --> [proposito].
```

```
de --> [de, este].
```

```
trabajo --> [trabajo].
```

```
copula --> [es].
```

```
verbo_proposito --> [analizar]
```

DCG para Movimiento3

```
movimiento3(mv3(F,P,C,V)) --> frase_de_proposito(F),  
                             pronombre(P), copula(C), verbo_proposito(V).
```

```
frase_de_proposito(fdp(A,P,D,T)) --> articulo(A),  
proposito(P), de(D), trabajo(T).
```

```
pronombre(pnom(nuestro)) --> [nuestro].
```

```
proposito(pro(proposito)) --> [proposito].
```

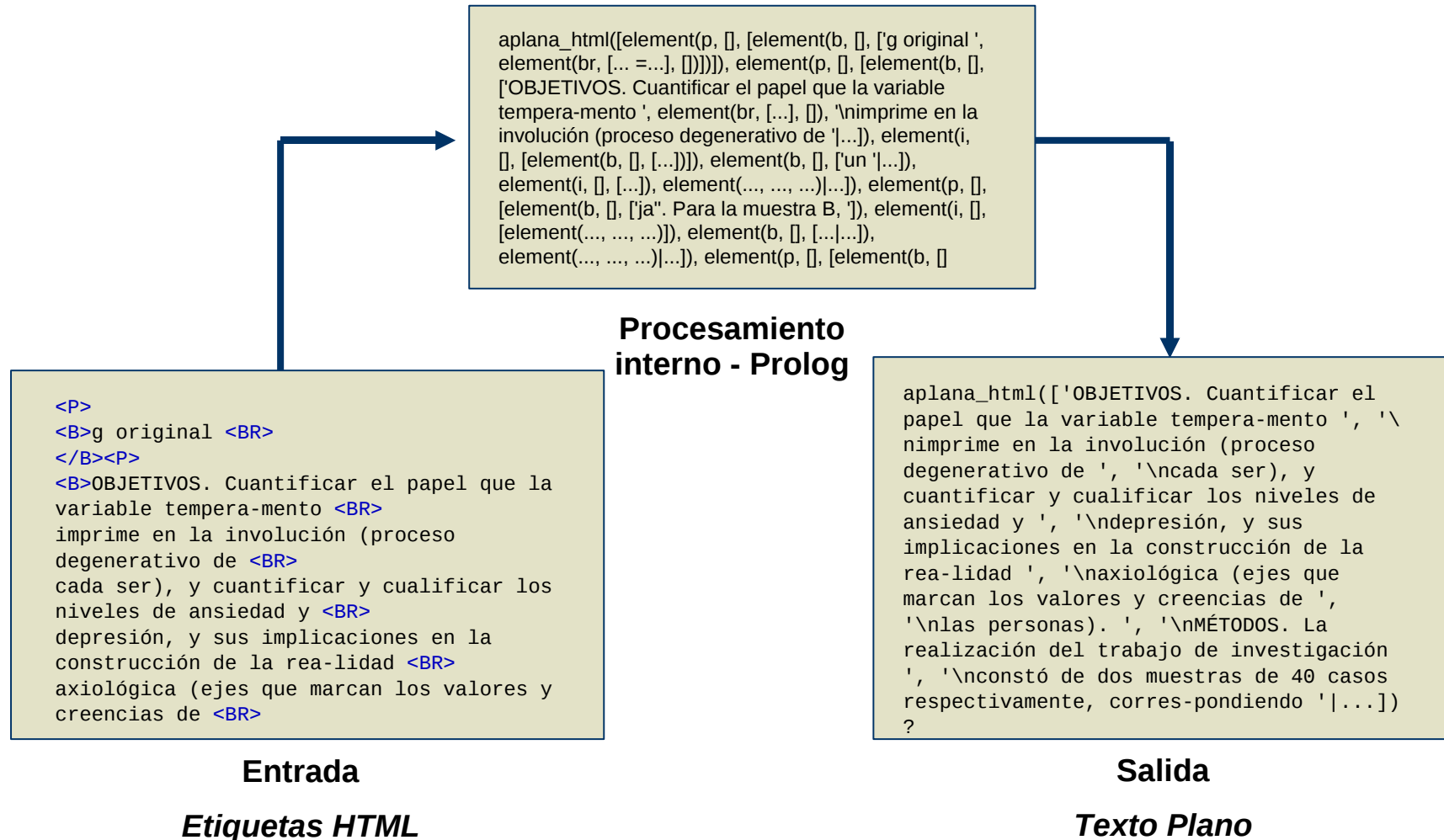
```
copula(cop(es)) --> [es].
```

```
verbo_proposito(verb(desarrollar)) --> [desarrollar].
```

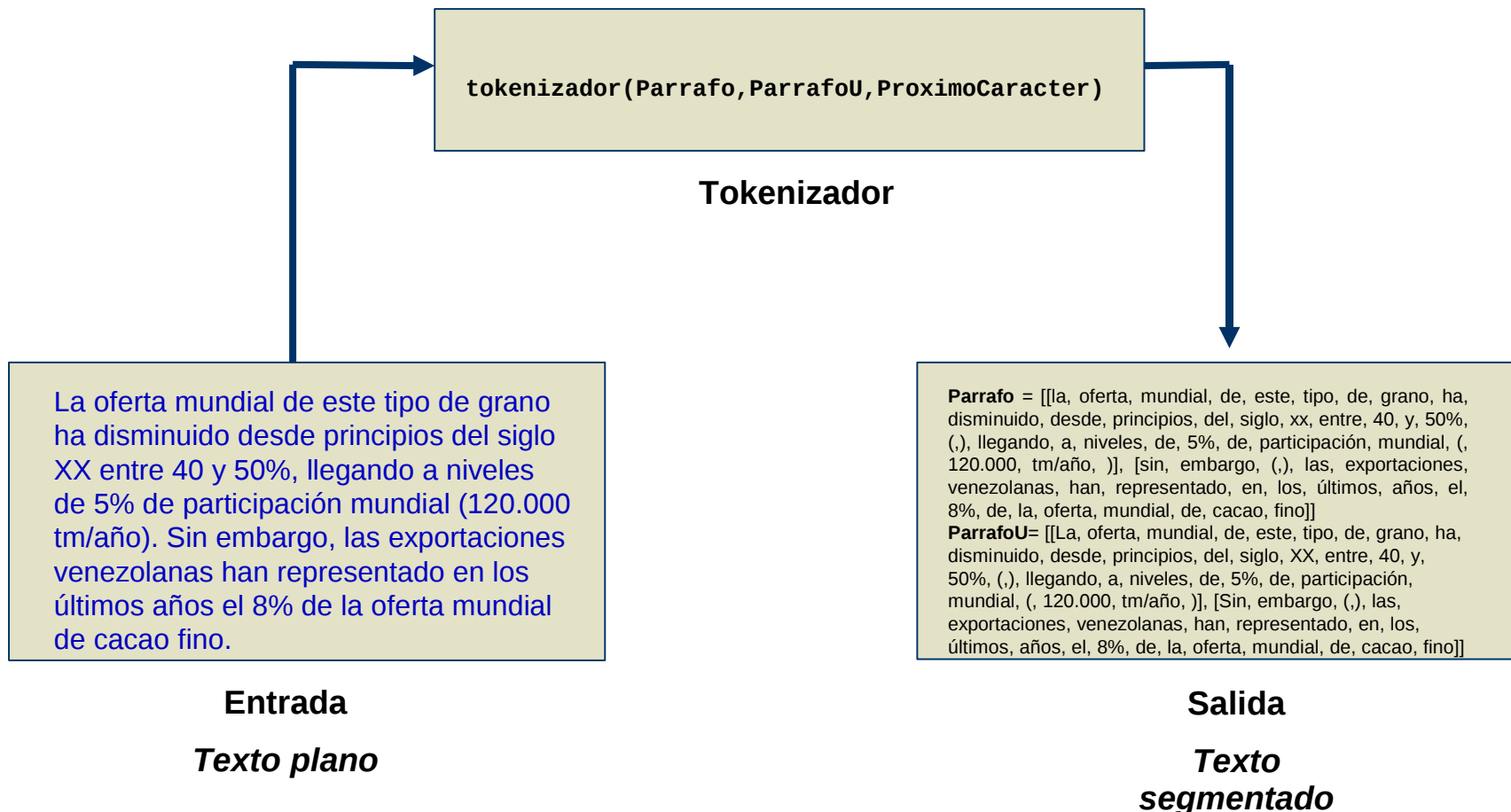
Resumidor: Componentes

Entrada	Componente	Salida
Texto	Tokenizador	Texto segmentado
Texto segmentado	Gramática	Texto etiquetado
Texto etiquetado	Claridad	Lista de tópicos [1]
Lista de tópicos [1]	Cohesión/ Coherencia	Lista de tópicos [2]
Lista de tópicos [2] + factor de resumen	Tópico común	Lista de tópicos [3]
Lista de tópicos [3] + Texto segmentado	Salida	Texto Resumen

Módulo: obtiene_artículo

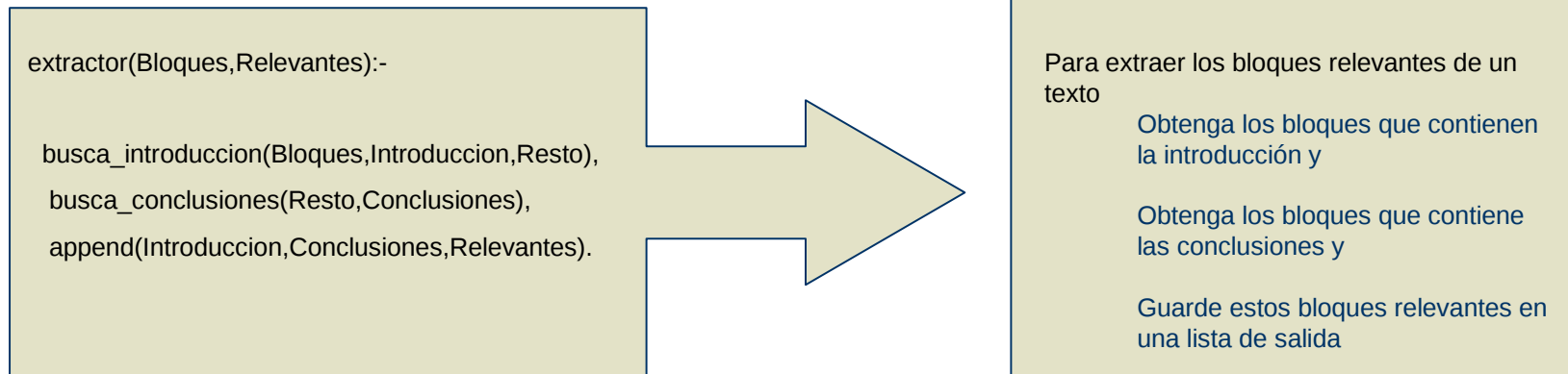


Microestructura Nivel-1



Módulo - Extractor

- Procesamiento de bloques relevantes



Módulo – Extractor
busca-introduccion

Identificación de Títulos

– Notación lingüística:

- Gramática de Tokens

```
introduccion --> p_introduccion, gap.  
introduccion --> gap, p_introduccion, gap.  
introduccion --> gap, p_introduccion.  
introduccion --> p_introduccion.  
  
p_introduccion --> [introduccion].  
p_introduccion --> [introducción].
```

■ Conceptualización:

- Corpus de estudio

```
titulo(Titulo) :-  
    name(Titulo, Lista_Ascii),  
    length(Lista_Ascii, N), N < 81,  
    % los titulos tienen menos de 80 caracteres.  
    tokenizer(Lista_Ascii, _), !.
```

■ Notación formal:

- Explotación de HTML

```
nuevo_titulo('<b>', '<font>').  
nuevo_titulo('<font>', '<b>').  
nuevo_titulo('<u>', '<b>').  
nuevo_titulo('<b>', '<u>').
```

Automatización: movimiento3

```
?- movimiento3(A,[el,proposito,de,este,estudio,es,investigar],L), a(A,0).
```

```
|----mv3
```

```
  |----fdp
```

```
    |----art
```

```
      |----el
```

```
        |----pro
```

```
          |----proposito
```

```
            |----d
```

```
              |----de
```

```
                |----este
```

```
                  |----trab
```

```
                    |----estudio
```

```
A = mv3(fdp(art(el), pro(proposito), d(de, este), trab(estudio)))
```

```
L = [es, investigar]
```

Procesamiento de Bloques Relevantes

- Invocación a las cláusulas del Extractor:

```
% ----- extractor
% extrae los bloques relevantes para el resumen de un documento dado.

extractor(Bloques,Previo, Relevantes):-
    busca_introduccion(Bloques,Previo,Introduccion,Resto),
    busca_conclusiones(Resto,Conclusiones, _),
    append(Introduccion,Conclusiones,Relevantes).
```

Módulo busca_introduccion: Fase de extracción

- Sub-programa que procesa el *contenido* del bloque propiamente dicho
- Almacenamos el *contenido* en “Resto” en el programa PROLOG

```
extrae_resto_introduccion([], [], []).  
extrae_resto_introduccion([Bloque1, Bloque2|RestoB], [], RestoB):-  
    nuevo_titulo(Bloque1, Bloque2), !.
```

```
extrae_resto_introduccion([B|Resto], [B|RI], RestoB):-  
    extrae_resto_introduccion(Resto, RI, RestoB).
```

Evaluando el Resumidor

<1> Se propone una metodología para obtener un Modelo de Gestión en una institución pública. <2> El modelo está basado en los procesos que intervienen en la Institución. <3> La metodología propuesta se divide en varias fases de trabajo y está basada en una matriz que considera los diferentes requerimientos que llegan a la institución y las unidades que intervienen en las actividades ofrecidas por la institución. <4> La primera fase levanta la información sobre la organización; la siguiente fase redefine aspectos claves de la organización, como son misión y objetivos. <5> La tercera fase describe los procesos implícitos en el modelo de gestión, para pasar a describir los requerimientos necesarios para su implementación. <6> Finalmente, está la fase de implantación del modelo en la organización. <7> Existen tres niveles de funcionamiento de la institución: dirección, servicios operativos y servicios de apoyo. <8> La metodología propuesta ha sido implementada en la Fundación para el Desarrollo de Ciencia y Tecnología en el Estado Mérida (FUNDACITE-Mérida) tomando en cuenta los lineamientos establecidos por el Ministerio de Ciencia y Tecnología. <9> La metodología propuesta es general, por lo que puede ser usada por cualquier organización y de manera repetida por una misma organización, cada vez que las condiciones de su entorno y de ella misma lo ameriten.

Evaluando el Resumidor

 <P>

<1> El tema de diseñar modelos de gestión para las organizaciones se ha justificado en la historia de la administración y la gerencia . <2> Normalmente, estos modelos se basan en un organigrama que presenta diversos niveles jerárquicos . <3> (1997) la gestión es la principal dimensión de una organización, en ella debe existir una necesidad permanente de interpretar el entorno, proyectar los cambios en él, y con un modelo de gestión estratégico poder dar respuesta a la realidad que exige dicho entorno. <4> Dezerega (1995) señala que el ideal gerencial es disponer de modelos para pre – evaluar y post – evaluar la viabilidad de planes para alcanzar objetivos y metas. <5> Ellos señalan que no sólo para cambiar es necesario el modelo, también es importante definirlo con claridad para que la organización tenga un estilo de trabajo en donde las fuerzas de todos se integren. <6> Es importante señalar lo difícil que es pensar en un modelo único de gestión para todas las organizaciones, es mejor pensar en un mecanismo de generación de modelos de gestión. <7> La metodología propuesta ha sido aplicada a la Fundación para el Desarrollo de la Ciencia y la Tecnología en el Estado Mérida (FUNDACITE-Mérida), la cual es una organización gubernamental sin fines de lucro, cuyo objetivo fundamental es la promoción de la Ciencia y la Tecnología en el Estado (Acta, 1990). <8> La metodología comienza con un diagnóstico de la institución . <9> Luego, se plantean los nuevos objetivos , misión, visión y estrategias para conseguirlos alineados a las líneas de acción del Ministerio de Ciencia y Tecnología (en nuestro caso, al Plan Nacional de Ciencia y Tecnología definido por ellos). <10> A continuación se describen los procesos y las normas que deben regir a la institución. <11> Las expectativas esperadas por la implantación de la misma en una organización , está orientado en hacer más pertinente las acciones de la misma con su entorno, haciendo más eficiente sus actividades y garantizando que sus actividades tengan un impacto real en el ambiente en que se mueve. <12> En el caso concreto de FUNDACITE-Mérida , la implantación del nuevo modelo de gestión propuesto por la metodología deberán convertir a la institución en un ente vinculador, gestor y concertador de actores y acciones en Ciencia y Tecnología más que en ejecutores, todo esto muy a tono con las conclusiones que se fueron obteniendo durante las fases de la metodología propuesta.

</P>

Evaluando el Resumidor

Estructura del resumen del autor	Estructura del resumen automático
<ol style="list-style-type: none">1. OBJETIVO DEL TRABAJO (UN MODELO DE GESTION).2. MODELO DE LA ORGANIZACIÓN.3. METODOLOGIA DE TRABAJO.4. FASES DE ESA METODOLOGIA DE TRABAJO.5. ORGANIZACIÓN donde se realiza el estudio.6. CUAL ES LA ORGANIZACIÓN en donde se “implementa” (aplica?) la metodología en cuestión.7. la METODOLOGIA ES GENERAL Y MODIFICABLE para otros contexto y organizaciones.	<ol style="list-style-type: none">1. OBJETIVO DEL TRABAJO: MODELOS DE GESTION EN ORGANIZACIONES.2. GESTION y MODELOS.3. IMPORTANCIA DE LOS MODELOS PARA LA GESTION.4. VINCULO ENTRE LOS MODELOS Y LA METODOLOGIA.5. CUAL ES LA ORGANIZACIÓN en donde se aplica (esta vez sin ambigüedades) la metodología en cuestión.6. COMIENZA LA METODOLOGIA.7. FASES DE LA METODOLOGIA.8. JUSTIFICA LA APLICACIÓN DE LA METODOLOGIA.9. APORTE DEL ESTUDIO.

Hacia donde

TEXTO 1: El presidente Ernesto Zedillo hizo un llamado a los partidos políticos para que el proceso electoral no sea utilizado como excusa para promover la división entre la población. (El Nacional, domingo 2 de agosto de 1998).

TEXTO 2: El presidente Ernesto Zedillo exhortó a los partidos políticos para que el proceso electoral no les excuse a promover la división entre la población