

# MÓDULO F1 • Fundamentos:

## Accesibilidad Generativa

"Si un LLM no puede leerte, no puede recomendarte."

### Objetivo del módulo

Asegurar que el contenido web está preparado técnica y estructuralmente para ser rastreado, comprendido e indexado por motores de generación de texto.

### 1. ¿Por qué es crítico este módulo?

Los modelos generativos acceden a contenido de dos formas principales:

1. **Durante su entrenamiento:** Contenido público de alta calidad que fue indexado o extraído mediante scraping durante la fase de entrenamiento del modelo.
2. **Durante la generación en tiempo real:** Cuando utilizan sistemas RAG (Retrieval-Augmented Generation) como Perplexity o Bing + ChatGPT para complementar sus respuestas con información actualizada.

En ambos casos, si tu web no está accesible, estructurada o autorizada correctamente, no formarás parte del universo informativo del modelo, por mucho que tu contenido sea valioso. Esto significa que serás invisible para los usuarios que buscan información a través de estos sistemas.

### Impacto en la visibilidad digital

La accesibilidad generativa es el primer filtro y el más crítico. Sin ella:

- Tu contenido no será considerado como fuente de información
- Tus competidores que sí implementen estas prácticas ganarán ventaja
- Perderás oportunidades de tráfico, autoridad y conversiones provenientes de LLMs

## 2. Componentes clave del módulo

### 2.1 • Indexabilidad y acceso técnico

**Acciones esenciales:**

#### **Configuración del archivo robots.txt**

El archivo robots.txt es la primera barrera que encuentran los rastreadores de IA. Debes asegurarte de permitir explícitamente el acceso a los bots relevantes:

```
User-agent: GPTBot  
Allow: /
```

```
User-agent: Google-Extended  
Allow: /
```

```
User-agent: CCBot  
Allow: /
```

```
User-agent: anthropic-ai  
Allow: /
```

**Nota importante:** Si deseas bloquear selectivamente algunas secciones de tu sitio (como páginas de checkout, áreas privadas o contenido sensible), puedes usar directivas específicas:

```
User-agent: GPTBot  
Allow: /blog/  
Allow: /recursos/  
Disallow: /checkout/  
Disallow: /mi-cuenta/
```

#### **Guía paso a paso para configurar tu robots.txt:**

1. **Localiza tu archivo robots.txt actual:**
2. Accede a tu sitio web añadiendo "/robots.txt" al final de tu dominio (ejemplo: tudominio.com/robots.txt)
3. Si no existe, deberás crearlo en la raíz de tu sitio web
4. **Edita el archivo:**
5. Si usas WordPress: Puedes usar plugins como Yoast SEO o Rank Math
6. Si tienes acceso FTP: Edita directamente el archivo en la raíz

7. Si usas otro CMS: Busca en la configuración SEO o consulta la documentación

**8. Añade los bots de IA:**

9. Copia el código de ejemplo anterior

10. Asegúrate de no eliminar otras directivas importantes que ya existan

11. Guarda los cambios

**12. Verifica la configuración:**

13. Usa herramientas como [robotstxt.org/robotstxt.html](https://robotstxt.org/robotstxt.html) para comprobar que la sintaxis es correcta

14. Visita nuevamente tu URL /robots.txt para confirmar que los cambios se han aplicado

**Herramientas adicionales para mejorar la indexabilidad:**

- **Bing Webmaster Tools:** Registra tu sitio para mejorar la visibilidad en Bing AI (que alimenta a ChatGPT con navegación)
- Paso 1: Crea una cuenta en [Bing Webmaster Tools](https://www.bing.com/webmaster)
- Paso 2: Verifica la propiedad de tu sitio (mediante archivo HTML, meta tag o DNS)
- Paso 3: Envía tu sitemap XML
- **IndexNow:** Notifica actualizaciones de contenido a múltiples motores de búsqueda simultáneamente
- Paso 1: Genera una clave API en [IndexNow](https://indexnow.io/)
- Paso 2: Crea un archivo con esa clave en la raíz de tu sitio
- Paso 3: Notifica URLs nuevas o actualizadas mediante la API

**Resultado esperado:** Los rastreadores de modelos generativos pueden acceder a tu contenido sin bloqueos, permitiendo que tu información sea considerada como fuente potencial.

## 2.2 • Estructura de contenidos para comprensión semántica

### Elementos críticos:

#### HTML semántico correcto

Los LLMs comprenden mejor el contenido cuando está estructurado con etiquetas HTML semánticas apropiadas:

```
<!-- Estructura incorrecta -->
<div class="titulo">Cómo optimizar tu web para IA generativa</div>
<div class="subtitulo">Pasos fundamentales</div>
<div class="texto">Este es el contenido principal...</div>

<!-- Estructura correcta -->
<h1>Cómo optimizar tu web para IA generativa</h1>
<h2>Pasos fundamentales</h2>
<p>Este es el contenido principal...</p>
<ul>
  <li>Primer paso importante</li>
  <li>Segundo paso importante</li>
</ul>
```

## Guía para implementar HTML semántico:

1. **Utiliza etiquetas de encabezado jerárquicas:**
  2. `<h1>` para el título principal (solo uno por página)
  3. `<h2>` para secciones principales
  4. `<h3>` para subsecciones
  5. Y así sucesivamente
6. **Emplea etiquetas específicas para cada tipo de contenido:**
  7. `<p>` para párrafos
  8. `<ul>` o `<ol>` para listas
  9. `<table>` para datos tabulares
  10. `<blockquote>` para citas
  11. `<figure>` y `<figcaption>` para imágenes con descripción
12. **Utiliza etiquetas semánticas modernas:**
  13. `<article>` para contenido independiente
  14. `<section>` para agrupar contenido relacionado
  15. `<nav>` para menús de navegación
  16. `<header>` y `<footer>` para encabezados y pies de página
  17. `<aside>` para contenido relacionado pero no esencial

## Implementación de datos estructurados (schema.org)

Los datos estructurados proporcionan contexto adicional a los LLMs sobre el tipo de contenido que están procesando:

## Ejemplo para una página de FAQ:

```

<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "FAQPage",
  "mainEntity": [{
    "@type": "Question",
    "name": "¿Qué es la optimización GEO?",
    "acceptedAnswer": {
      "@type": "Answer",
      "text": "GEO (Generative Engine Optimization) es el conjunto de técnicas para
optimizar contenido web para ser comprendido, citado y recomendado por
modelos de lenguaje como ChatGPT, Claude o Perplexity."
    }
  },
  {
    "@type": "Question",
    "name": "¿En qué se diferencia GEO del SEO tradicional?",
    "acceptedAnswer": {
      "@type": "Answer",
      "text": "Mientras que el SEO tradicional se enfoca en posicionar palabras clave
en motores de búsqueda, GEO se centra en la comprensión semántica, la
estructura conversacional y la autoridad para que los modelos generativos citen tu
contenido."
    }
  }
]}
</script>

```

### Tipos de schema.org más relevantes para GEO:

- **FAQPage:** Ideal para páginas de preguntas frecuentes
- **HowTo:** Perfecto para tutoriales y guías paso a paso
- **Article:** Para entradas de blog y artículos informativos
- **Product:** Para páginas de productos
- **LocalBusiness:** Para negocios con presencia física
- **Person:** Para perfiles de autor o experto
- **Organization:** Para información sobre tu empresa

### Guía paso a paso para implementar schema.org:

1. **Identifica el tipo de contenido** que quieres marcar
2. **Genera el código JSON-LD** utilizando:
3. [Schema Markup Generator](#)
4. [Google's Structured Data Markup Helper](#)
5. **Inserta el código** en la sección `<head>` de tu página HTML
6. **Valida tu implementación** con la [herramienta de prueba de datos estructurados de Google](#)

**Resultado esperado:** El contenido se presenta de forma estructurada, clara y clasificable por los LLMs, facilitando su comprensión y aumentando la probabilidad de que sea citado.

## 2.3 • Metadatos claros y actualizados

Los metadatos son cruciales para que los LLMs entiendan la frescura, relevancia y contexto de tu contenido:

### Elementos esenciales:

- **Fecha visible de publicación y última revisión:** `` `html

Publicado: 15 de marzo de 2025

Última actualización: 1 de junio de 2025

`` `

- **Autor claramente identificado:** `` `html

Autor: [María Rodríguez](#), Especialista en Marketing Digital

`` `

- **Afiliaciones y fuentes:** `` `html

Fuentes: [Estudio de Caso XYZ](#), [Investigación ABC](#)

`` `

### Guía para optimizar metadatos:

1. **Implementa fechas en formato estándar:**
2. Usa el atributo `datetime` con formato ISO (YYYY-MM-DD)
3. Muestra la fecha en formato legible para humanos
4. Actualiza la fecha de revisión cuando hagas cambios significativos
5. **Crea una estructura consistente de información de autor:**
6. Incluye nombre completo
7. Añade credenciales relevantes o cargo
8. Enlaza a una página de perfil o LinkedIn
9. **Implementa metadatos técnicos:**

10. Meta título: `<title>Título optimizado para GEO | Nombre de marca</title>`

11. Meta descripción:

`<meta name="description" content="Descripción conversacional que responde a una pregunta específica...">`

12. Open Graph para compartir en redes sociales: `html <meta property="og:title"`

`content="Título optimizado"> <meta property="og:description"`

`content="Descripción conversacional"> <meta property="og:image"`

`content="URL-de-imagen"> <meta property="og:url" content="URL-canónica">`

**Resultado esperado:** El contenido tiene señales claras de frescura, trazabilidad y confiabilidad, valores fundamentales para los sistemas generativos modernos.

## 2.4 • Autoría y autoridad básica

La identificación clara de la fuente es esencial para que los LLMs consideren tu contenido como confiable:

### Elementos clave:

- **Firma visible de autor:** ````html`

María Rodríguez

### María Rodríguez

Especialista en Marketing Digital con 10 años de experiencia en optimización de contenidos.

[LinkedIn](#) | [Twitter](#)

```

- **Página "Sobre nosotros" completa:**

- Historia de la empresa/proyecto
- Misión y valores
- Equipo con perfiles detallados
- Logros y reconocimientos

- Clientes o casos de éxito

- **Información de contacto real:**

- Dirección física (si aplica)
- Correo electrónico de contacto
- Número de teléfono (si aplica)

- Formulario de contacto funcional
- Perfiles de redes sociales activos

### Guía para implementar señales de autoridad:

1. **Crea una página de autor para cada contribuidor:**
2. Incluye foto profesional
3. Detalla experiencia relevante
4. Lista publicaciones anteriores
5. Añade credenciales y certificaciones
6. Enlaza perfiles profesionales externos
7. **Desarrolla una página "Sobre nosotros" completa:**
8. Cuenta la historia de tu marca/proyecto
9. Explica tu misión y valores
10. Presenta al equipo con fotos y biografías
11. Incluye testimonios o casos de éxito
12. Añade logos de clientes o menciones en medios
13. **Implementa información de contacto visible:**
14. Añade un pie de página con datos básicos
15. Crea una página de contacto detallada
16. Incluye un mapa si tienes ubicación física
17. Ofrece múltiples canales de comunicación

**Resultado esperado:** La IA asocia tu sitio con una fuente real, confiable y con identidad verificada, aumentando la probabilidad de que te cite como referencia.

## 3. Criterios de evaluación del módulo F1

Utiliza esta tabla para evaluar tu implementación de los fundamentos de accesibilidad generativa:

| Criterio                | 0 = Ausente          | 1 = Débil                                     | 2 = Aceptable                       | 3 = Óptimo                                                         |
|-------------------------|----------------------|-----------------------------------------------|-------------------------------------|--------------------------------------------------------------------|
| robots.txt y bots de IA | No existe robots.txt | Existe pero sin configuración para bots de IA | Configurado para algunos bots de IA | Configuración completa y optimizada para todos los bots relevantes |



| Criterio                          | 0 = Ausente                         | 1 = Débil                                          | 2 = Aceptable                                  | 3 = Óptimo                                                           |
|-----------------------------------|-------------------------------------|----------------------------------------------------|------------------------------------------------|----------------------------------------------------------------------|
| Schema.org implementado           | Sin implementación                  | Implementación básica con errores                  | Implementación correcta pero limitada          | Implementación completa y validada para todos los tipos de contenido |
| HTML semántico y estructura       | Uso de divs genéricos sin semántica | Estructura básica con algunos elementos semánticos | Buena estructura semántica con jerarquía clara | Estructura semántica perfecta con todos los elementos apropiados     |
| Metadatos y fecha de revisión     | Sin metadatos visibles              | Metadatos básicos sin fechas                       | Metadatos completos con fecha de publicación   | Metadatos completos con fechas de publicación y actualización        |
| Firma de autor y entidad visible  | Sin información de autor            | Nombre de autor sin detalles                       | Perfil básico de autor con algunos detalles    | Perfil completo con credenciales, foto y enlaces                     |
| Herramientas de indexación usadas | Ninguna herramienta                 | Una herramienta básica                             | Varias herramientas sin seguimiento            | Suite completa de herramientas con monitoreo activo                  |

**Puntuación total posible: 18 puntos**

## 4. Recursos complementarios

Para profundizar en la implementación de este módulo, te recomendamos estos recursos:

- [Lista actualizada de bots de IA que rastrean sitios](#)
- [Guía completa de datos estructurados de Google](#)
- [Verificador de robots.txt](#)
- [Herramienta de validación de Schema.org](#)

- [Guía de IndexNow para notificación de contenido](#)

## 5. Output esperado – Módulo F1: Accesibilidad Generativa

Al finalizar este módulo, tu contenido estará técnicamente preparado para ser leído, interpretado y utilizado por los principales motores de generación (como ChatGPT, Claude o Perplexity).

Tu sitio web permitirá el acceso de crawlers útiles mediante robots.txt correctamente configurado

La información clave estará estructurada con schema.org (datos enriquecidos) para máxima comprensión

La autoría, fechas y contexto editorial estarán claramente visibles, aumentando la confiabilidad

Las URLs y encabezados tendrán una jerarquía semántica lógica que facilita la navegación y comprensión

No habrá errores de indexación ni duplicidades técnicas que confundan a los sistemas de IA

**Resultado práctico:** Tu contenido pasa de ser "visible en buscadores" a ser comprensible y usable por modelos generativos, el primer paso fundamental para aparecer en sus respuestas.

## 6. Checklist de implementación

Utiliza esta lista para asegurarte de haber implementado todos los elementos clave:

- ☐ Archivo robots.txt configurado para permitir bots de IA relevantes
- ☐ HTML semántico implementado en todas las páginas importantes
- ☐ Schema.org añadido para el tipo de contenido principal
- ☐ Fechas de publicación y actualización visibles
- ☐ Información de autor completa y enlazada
- ☐ Página "Sobre nosotros" detallada y profesional
- ☐ Información de contacto visible y funcional
- ☐ Sitio registrado en Bing Webmaster Tools
- ☐ Sistema IndexNow configurado para notificaciones
- ☐ Validación técnica completada sin errores

## 7. Prompt para implementación con LLMs

Puedes utilizar el siguiente prompt con ChatGPT, Claude u otro LLM para ayudarte a implementar los aspectos técnicos de este módulo:

Actúa como un consultor experto en GEO (Generative Engine Optimization). Necesito implementar las bases técnicas para que mi sitio web sea accesible para modelos de IA generativa.

Mi sitio web es: [DESCRIBE TU SITIO, TEMÁTICA Y PLATAFORMA]

Por favor, ayúdame a:

1. Crear un archivo robots.txt optimizado para permitir el acceso a bots de IA relevantes
2. Generar el código JSON-LD de schema.org apropiado para mi tipo de contenido principal
3. Revisar y mejorar la estructura HTML semántica de mis páginas
4. Implementar correctamente los metadatos y la información de autor

Necesito código específico y pasos detallados que pueda implementar directamente.

---

**Próximo módulo:** F2 • Contexto semántico y formato óptimo - Donde aprenderás a estructurar tu contenido para que los LLMs no solo lo lean, sino que lo entiendan y valoren como fuente relevante.