

# Spectrum Technology Platform

Version 12.0

Guía de extracción de información



# Contents

## 1 - Introducción

---

Módulo de Extracción de información	4
Compatibilidad de idiomas	4
Seguridad del modelo	5

## 2 - Categorización de texto

---

Introducción a la Categorización de texto	7
Preparación de los datos	7
Opciones de configuración	8
Capacitación del modelo	12
Evaluación del modelo	12
Categorización de texto	12

## 3 - Extracción de entidad

---

Introducción	15
Entidades preexistentes	15
Entidades personalizadas	16

## 4 - Extracción de relaciones

---

Introducción	24
Tipos de relación	25

## 5 - Comandos de la Utilidad de administración

---

Comandos de la Utilidad de administración	30
iemodel delete	30
iemodel evaluate model	30

iemodel evaluate train_model	33
iemodel export	35
iemodel import	35
iemodel list	36
iemodel train	37
iemodel trainAndevaluate model	37

## 6 - Referencia de etapas

---

Componentes de Information Extraction	42
Read from Documents	42
Entity Extractor	47
Text Categorizer	50
Extractor de relaciones	52

# 1 - Introducción

## In this section

---

Módulo de Extracción de información	4
Compatibilidad de idiomas	4
Seguridad del modelo	5

## Módulo de Extracción de información

El módulo Extracción de información ofrece capacidades de perfeccionamiento de texto y extracción de información avanzadas de texto de entrada en lenguaje natural.

### Características proporcionadas

<b>Categorización de texto</b>	Le permite asignar categorías personalizadas a texto no estructurado. Esto es posible luego de haber capacitado un <i>modelo de categorización de texto</i> mediante el uso de la utilidad de administración. Esta función se puede utilizar para indexar los informes de atención de pacientes, clasificar documentos por dominio y subdominio, y categorizar el correo electrónico como correo no deseado y correo deseado, entre otras aplicaciones.
<b>Extracción de entidad</b>	Le permite capacitar modelos para extraer entidades a partir de datos no estructurados. El módulo incluye algunas <i>entidades preexistentes</i> .  Si se requiere, defina entidades personalizadas mediante el uso del tipo de modelo <i>CustomEntity</i> . Después de crear y capacitar modelos específicos de dominio, puede extraer entidades según el modelo que haya capacitado.
<b>Extracción de relaciones</b>	Le permite identificar el tipo de relación que relaciona a un par de entidades en cualquier texto de entrada en lenguaje natural.

## Compatibilidad de idiomas

En todas las etapas del módulo Extracción de información, la versión actual solo admite funciones de extracción de información para texto de entrada en idioma *inglés*.

**Nota:** En el caso de la etapa **Entity Extractor**, además del inglés, se admiten los siguientes idiomas en la fase *beta*:

<b>es</b>	Español (México)
<b>fr</b>	Francés
<b>de</b>	Alemán
<b>pt</b>	Portugués (Brasil)

**Importante:** Los idiomas *beta* solo se encuentran disponibles en el caso de *Custom Entity* (Entidades personalizadas), y no para entidades preexistentes.

## Seguridad del modelo

Se deben asignar permisos de seguridad en **Management Console** para realizar varias funciones con Extracción de información:

- Se requieren permisos de visualización para categorizar o enumerar el modelo.
- Se requieren permisos de modificación para volver a capacitar o importar el modelo (si el modelo ya existe).
- Se requieren permisos de creación para importar o capacitar el modelo.
- Se requieren permisos de eliminación para borrar el modelo.

# 2 - Categorización de texto

## In this section

---

Introducción a la Categorización de texto	7
Preparación de los datos	7
Opciones de configuración	8
Capacitación del modelo	12
Evaluación del modelo	12
Categorización de texto	12

## Introducción a la Categorización de texto

La categorización de texto, también conocida como clasificación de texto, es el proceso de asignación de categorías personalizadas al contenido no estructurado o texto sin formato, como correos electrónicos, artículos noticiosos y comentarios, según cuánto de dicho contenido coincide con la categoría. La categorización se puede categorizar por tema, autor, fecha o casi cualquier sistema de clasificación que se conciba.

Para crear su propio categorizador, debe capacitar un modelo de categorizador con sus datos y categorías. El capacitador analiza los datos y guarda la información que obtiene en el proceso de capacitación. Luego analiza el contenido y determina la categoría a la cual pertenece el contenido.

La función de categorización de texto utiliza un proceso de categorización estadística del texto. Esto aplica métodos de aprendizaje de máquina para aprender reglas de clasificación automática basadas en documentos de capacitación rotulados por humanos.

Dado que es posible aplicar la categorización que elija, primero debe "capacitar" al modelo para que "aprenda" las categorías. Luego de esto, puede usar ese modelo en la etapa **Text Categorizer** para categorizar los datos no estructurados.

Spectrum™ Technology Platform utiliza los comandos de la utilidad de administración para administrar modelos de categorización de texto. Para obtener una descripción de estos comandos, consulte [Comandos de la Utilidad de administración](#) en la página 30.

## Preparación de los datos

El primer paso para utilizar la categorización de texto es preparar el archivo de entrada y el archivo de prueba. Para ello, debe estructurar los datos como valores separados por tabuladores en ambos archivos. Los archivos deben tener detalles en este formato:

- Codificación UTF-8
- Datos separados por tabuladores en dos columnas, donde la primera columna contiene el nombre de categoría (por ejemplo: "Paciente" o "Proveedor") y la segunda columna tiene los datos de cada categoría (como se muestra en el ejemplo a continuación)

Los datos deben verse así:

```
Patient      John Smith dob04181963 224 Main St. Atl GA 30311
Provider     Mark Johnson M.D. NPI5489512047 412 Washington Atl GA 30301
```

## Opciones de configuración

Esto involucra la creación de un archivo de `Opciones de capacitación` que contiene información sobre su modelo y las opciones que puede aplicar para la capacitación del modelo. Este archivo debe estar en formato XML con codificación UTF-8 y debe incluir este encabezado y las características de capacitación requeridas:

### *Encabezado en el archivo Opciones de capacitación*

El encabezado menciona detalles del modelo, su tipo y la ruta de la prueba y los archivos de entrada.

- `modelName`: nombre del modelo
- `modelType`: el tipo del modelo (que es `TC`, lo que significa categorización de texto en este caso)
- `modelDescription`: descripción del modelo
- `inputFilePath`: ubicación del archivo de entrada utilizado para capacitar el modelo
- `testFilePath`: ubicación del archivo de prueba

#### **Nota:**

El archivo de prueba mide la eficacia de un modelo. Este determina el comportamiento del modelo personalizado con diversos parámetros de capacitación. Como una buena práctica, debe usar distintos archivos de entrada y de prueba en la capacitación o la evaluación de sus modelos personalizados.

`algoritmo`: el algoritmo de Machine Learning utilizado para capacitar el modelo (el valor predeterminado es `MaxEnt`)

### *Características de capacitación*

Estas son características de capacitación que puede usar para crear una nueva categoría.

**Nota:** Si utiliza varias características, las puede ubicar en cualquier orden dentro del archivo.

- **Característica lingüística:** para especificar las propiedades de idioma
  - `Stemming`: reduce las palabras a su raíz. Por ejemplo, "aseguradora", "asegurado" y "asegurar" se pueden reducir a la raíz "seguro".

```
<trainingFeature>
  <featureName>Stemming</featureName>
</trainingFeature>
```

- **Características de palabras clave:** para definir la lista de palabras clave
  - `IgnoreWords`: también denominadas palabras irrelevantes, esta función filtra las palabras comunes que no afectan la categorización, como "el/la", "y", "pero". Estas palabras se deben

separar solo por una coma, sin espacios. También puede usar la clave `Adjuntar` con esta característica, que cuando la define en "Verdadero", se agrega a la lista existente de palabras irrelevantes.

```
<trainingFeature>
  <featureName>IgnoreWords</featureName>
  <featureParams>
    <entry>
      <key>WordList</key>
      <value>
        and, the, for, with, still, tri, rep, cust, keep, get, req, call
      </value>
    </entry>
    <entry>
      <key>Append</key>
      <value>True</value>
    </entry>
  </featureParams>
</trainingFeature>
```

- `CategoryKeywords`: identifica una categoría para una lista de palabras clave que pertenecen a múltiples listas personalizadas. Por ejemplo, Días de semana en la lista `CategoryKeywords` contiene Palabras clave como Lunes, Martes, Miércoles, Jueves y Viernes.

Puede especificar esta característica en forma opcional si el cruce debe distinguir entre mayúsculas y minúsculas. Cuando la usa, el valor predeterminado es `true`.

```
<trainingFeature>
  <featureName>CategoryKeywords</featureName>
  <featureParams>
    <entry>
      <key>Weekdays</key>
      <!-- List of weekdays -->
      <value>Monday, Tuesday, Wednesday, Thursday, Friday</value>
    </entry>
    <entry>
      <key>WeekendDays</key>
      <!-- List of weekend days -->
      <value>Saturday, Sunday</value>
    </entry>
    <entry>
      <key>CaseSensitive</key>
      <value>True</value>
    </entry>
  </featureParams>
</trainingFeature>
```

- **KeyWords:** busca las palabras que especificó como pertenecientes a una lista personalizada, como *DaysOfWeek* o *Month*. De forma opcional, también especifica si el cruce debe distinguir mayúsculas de minúsculas; cuando se utiliza, el valor predeterminado es "verdadero".

```
<trainingFeature>
  <featureName>KeyWords</featureName>
  <featureParams>
    <entry>
      <key>KeyWordList</key>
      <value>Monday, Tuesday</value>
    </entry>
    <entry>
      <key>CaseSensitive</key>
      <value>False</value>
    </entry>
  </featureParams>
</trainingFeature>
```

- **Característica léxica:** para especificar las propiedades de lexema
  - **NGram:** busca parte de una cadena más larga, donde "n" representa el número de palabras que se buscarán. Por ejemplo, si buscara la frase "ser o no ser", podría buscar un unigrama de "ser" o "no", o un bigrama de "ser o" o "no ser", un trigram de "ser o no" o "o no ser", etc.

```
<trainingFeature>
  <featureName>NGram</featureName>
  <featureParams>
    <entry>
      <key>Count</key>
      <value>3</value>
    </entry>
  </featureParams>
</trainingFeature>
```

Un ejemplo de archivo de opciones de capacitación:

```
<trainingOptions>
  <modelName>modelone</modelName>
  <modelType>TC</modelType>
  <modelDescription>modelOne</modelDescription>

  <inputFilePath>C:/SpectrumIE/textclassification/train_Input.csv</inputFilePath>

  <testFilePath>C:/SpectrumIE/textclassification/train_Test.txt</testFilePath>

  <algorithm>SVM</algorithm>

  <trainingFeatures>
```

```

<!-- Keyword features -->
<trainingFeature>
  <featureName>IgnoreWords</featureName>
  <featureParams>
    <entry>
      <key>WordList</key>
      <value>
        and,the,for,with,still,tri,rep,cust,keep,get,req,call
      </value>
    </entry>
    <entry>
      <key>Append</key>
      <value>True</value>
    </entry>
  </featureParams>
</trainingFeature>

<trainingFeature>
  <featureName>CategoryKeywords</featureName>
  <featureParams>
    <entry>
      <key>Category1</key>
      <value>CategoryKeyword1,CategoryKeyword2</value>
    </entry>
    <entry>
      <key>Category2</key>
      <value>CategoryKeyword3,CategoryKeyword4</value>
    </entry>
  </featureParams>
</trainingFeature>

<trainingFeature>
  <featureName>KeyWords</featureName>
  <featureParams>
    <entry>
      <key>KeyWordList</key>
      <value>
        jam,misfeed,install,help,mechanical,failure,jam,pc,connection
      </value>
    </entry>
  </featureParams>
</trainingFeature>

<!-- Linguistic feature -->
<trainingFeature>
  <featureName>Stemming</featureName>
</trainingFeature>

<!-- Lexical feature -->
<trainingFeature>
  <featureName>NGram</featureName>
  <featureParams>
    <entry>

```

```

    <key>Count</key>
    <value>3</value>
  </entry>
</featureParams>
</trainingFeature>

</trainingFeatures>
</trainingOptions>

```

## Capacitación del modelo

Después de crear un archivo de opciones, debe capacitar su modelo para descubrir posibles relaciones predictivas. Para ello, aplique los métodos de aprendizaje automático. Spectrum™ Technology Platform utiliza el comando **iemodel train** en la página 37 CLI para capacitar un modelo. Después de capacitar el modelo, puede usarlo en la categorización.

## Evaluación del modelo

Podría querer probar el modelo después de capacitarlo para asegurarse de que el archivo de opciones de capacitación sea correcto y las categorías se estén asignando según lo esperado.

Se puede probar el modelo con el comando **iemodel trainAndevaluate model** en la página 37 CLI.

## Categorización de texto

1. Cree un flujo de datos que incluya una etapa de origen como **Read from File** o **Read from XML**, la etapa **Text Categorizer**, y una etapa receptora como **Write to File** o **Write to XML**.
2. Durante la etapa de origen, indique el archivo de entrada.
3. Durante la etapa **Text Categorizer**, seleccione el modelo en el campo **Nombre del categorizador**. Este es el modelo por el cual recibió capacitación en la fase de categorización de texto. Para obtener más información sobre la capacitación de un modelo, consulte [Capacitación del modelo](#) en la página 12
4. En el campo **Conteo de categorías**, seleccione la cantidad de niveles de comparación de la categoría que se deben incluir en los datos de salida. Por ejemplo, el cruce más cercano o el que le sigue en cercanía.

**Nota:** El valor máximo corresponde a la cantidad de categorías diferentes especificadas durante la capacitación del modelo.

5. Haga clic en **Aceptar**.
6. Ejecute el trabajo.

# 3 - Extracción de entidad

## In this section

---

Introducción	15
Entidades preexistentes	15
Entidades personalizadas	16

## Introducción

La extracción de entidades es el proceso de identificación y recuperación de entidades a partir de datos no estructurados. Puede utilizar las entidades preexistentes que forman parte de la etapa **Entity Extractor** o puede capacitar un modelo para extraer entidades personalizadas.

## Entidades preexistentes

Las entidades preexistentes son aquellas que vienen con el módulo **Information Extraction**.

Para ver una lista de las entidades preexistentes, abra la etapa **Entity Extractor**, seleccione la casilla de verificación **Anular opciones predeterminadas del sistema con los siguientes valores** y haga clic en **Agregado rápido**. La lista de las entidades se muestra en la sección **Seleccionar entidades**.

- *Person*
- *Address*
- *ProperNouns*
- *ISBN*
- *CreditCard*
- *ZipCode*
- *WebAddress*
- *Mention*
- *HashTag*
- *SSN*
- *Phone*
- *Email*
- *Date*
- *Location*
- *Organization*

Siga los pasos restantes de esta sección para extraer estos tipos de entidades de sus datos.

## Extracción de entidades preexistentes

1. Cree un flujo de datos que incluya una etapa de origen **Read from Documents**, una etapa **Entity Extractor** y una etapa receptora como **Write to File** o **Write to XML**.
2. Durante la etapa de origen, indique el archivo de entrada.
3. En la etapa **Entity Extractor**, seleccione las entidades según los datos que desea extraer del archivo de entrada. Por ejemplo, si desea seleccionar los nombres de todas las personas y direcciones del archivo, seleccione las entidades *Address* (Dirección) y *Person* (Persona).

**Nota:** *Address* y *Person* son las entidades predeterminadas. Para extraer datos según cualquier otra entidad, seleccione la casilla de verificación **Anular opciones predeterminadas del sistema con los siguientes valores** y haga clic en **Agregado rápido**. La lista de las entidades se muestra en la sección **Seleccionar entidades**.

4. Para obtener la frecuencia en el archivo de entrada de los datos relacionados con las entidades especificadas, seleccione la casilla de verificación **Conteo de entidades de salida**.
5. Haga clic en **Aceptar**.
6. Ejecute el trabajo.

## Entidades personalizadas

Como sucede con entidades preexistentes, también puede capacitar modelos para recuperar las entidades personalizadas. Estas entidades pueden pertenecer a cualquier dominio y ser de cualquier tipo. Por ejemplo, puede usar textos médicos para extraer una lista de diagnósticos o productos farmacéuticos. El proceso de extracción de entidades personalizadas incluye:

1. Preparación de datos: preparación del archivo de entrada y del archivo de prueba
2. Configuración de las opciones: creación de archivo de opciones de capacitación que contiene información sobre el modelo y las opciones a aplicar durante la capacitación del modelo
3. Capacitación del modelo
4. Extracción de entidades

Cuando se llevan a cabo correctamente todos estos pasos, el nuevo tipo de entidad se agrega a la lista en la etapa **Entity Extractor**, y lo puede usar para extraer detalles desde un archivo no estructurado.

## Preparación de datos para entidades personalizadas

El primer paso para crear entidades personalizadas es preparar el archivo de entrada y el archivo de prueba. La función de entidades personalizadas exige que las entidades de esos archivos estén rodeadas de MagicWord que usted especifica en el archivo de opciones de capacitación (que se analiza en el siguiente tema).

Supongamos que está extrayendo diagnósticos de datos no estructurados de su archivo de entrada y ha designado la MagicWord *DIAGNOSIS* (diagnóstico) en el archivo de opciones de capacitación. Cada vez que aparezca el nombre de una enfermedad o una condición en el texto, la palabra estará encerrada con esa MagicWord, como se ve a continuación:

```
The term diagnostic criteria designates the specific combination of
signs, symptoms, and test results that the clinician uses to attempt
to determine the correct diagnosis. Some examples of diagnostic
criteria, also known as clinical case definitions, are: Amsterdam
criteria for DIAGNOSIShereditary nonpolyposis colorectal cancerDIAGNOSIS
McDonald criteria for DIAGNOSISmultiple sclerosisDIAGNOSIS ACR criteria
for DIAGNOSISsystemic lupus erythematosusDIAGNOSIS Centor criteria for
DIAGNOSISstrep throatDIAGNOSIS.
```

Para obtener información sobre cómo identificar la MagicWord, consulte el tema siguiente.

## Configuración de las opciones para las entidades personalizadas

Esto involucra la creación de un archivo de `Opciones de capacitación` que contiene información sobre su modelo y las opciones que puede aplicar para la capacitación del modelo. Este archivo debe estar en formato XML con codificación UTF-8 y debe incluir este encabezado y las características de capacitación requeridas:

### *Encabezado en el archivo Opciones de capacitación*

El encabezado menciona detalles del modelo, la ruta de la prueba y los archivos de entrada, además de una palabra clave para anotar las entidades personalizadas.

- `modelName`: nombre del modelo personalizado
- `modelType`: tipo del modelo personalizado (que es *CustomEntity*).
- `modelDescription`: descripción del modelo personalizado
- `inputFilePath`: ruta del archivo etiquetado utilizado para capacitar el modelo (archivo de entrada)
- `testFilePath`: ruta del archivo utilizado para probar el modelo
- `magicWord`: palabra clave utilizada para anotar las entidades personalizadas

- idioma: el idioma utilizado en el texto.

**Nota:** Se admite el inglés. El holandés, el francés, el alemán y el español están en la etapa de desarrollo beta.

### Características de capacitación

Puede usar estas características de capacitación para crear las entidades personalizadas

- **Características lingüísticas:** para especificar las propiedades de idioma
  - `POSTagger`: etiqueta para identificar partes del discurso, como sustantivos, pronombres, adjetivos y verbos.

```
<trainingFeature>
  <featureName>POSTagger</featureName>
</trainingFeature>
```

- **Características ortográficas:** para especificar las propiedades estructurales
  - `CaseIdentifier`: identifica si las entidades personalizadas están completamente en mayúsculas, en minúsculas o en una combinación de ambas.

```
<trainingFeature>
  <featureName>CaseIdentifier</featureName>
</trainingFeature>
```

- `NumericIdentifier`: identifica si las entidades personalizadas son numéricas o alfanuméricas.

```
<trainingFeature>
  <featureName>NumericIdentifier</featureName>
</trainingFeature>
```

- `1st2ndIdentifier`: identifica si las entidades personalizadas son ordinales, como 1.<sup>a</sup>, 2.<sup>a</sup> y 3.<sup>a</sup>.

```
<trainingFeature>
  <featureName>1st2ndIdentifier</featureName>
</trainingFeature>
```

- `PatternMatcher`: compara palabras con uno o más patrones con expresiones regulares. Cuando se proporcionan múltiples expresiones, incluye la condición de combinación `AND` para todas las expresiones o `OR` (valor predeterminado) para cualquier expresión.

```
<trainingFeature>
  <featureName>PatternMatcher</featureName>
  <featureParams>
    <entry>
      <key>RegEx1</key>
```

```

    <value>b[aeiou]t</value>
  </entry>
  <entry>
    <key>RegEx2</key>
    <value>b[xyz]t</value>
  </entry>
  <entry>
    <key>JoinCondition</key>
    <value>AND</value>
  </entry>
</featureParams>
</trainingFeature>

```

- **Características de palabras clave:** para definir la lista de palabras clave

- **CategoryKeywords:** identifica una categoría para una lista de palabras clave que pertenecen a múltiples listas personalizadas. Por ejemplo, Días de semana en la lista `CategoryKeywords` contiene `Palabras clave` como `Lunes`, `Martes`, `Miércoles`, `Jueves` y `Viernes`.

Puede especificar esta característica en forma opcional si el cruce debe distinguir entre mayúsculas y minúsculas. Cuando la usa, el valor predeterminado es `true`.

```

<trainingFeature>
  <featureName>CategoryKeywords</featureName>
  <featureParams>
    <entry>
      <key>Weekdays</key>
      <!-- List of weekdays -->
      <value>Monday, Tuesday, Wednesday, Thursday, Friday</value>
    </entry>
    <entry>
      <key>WeekendDays</key>
      <!-- List of weekend days -->
      <value>Saturday, Sunday</value>
    </entry>
    <entry>
      <key>CaseSensitive</key>
      <value>True</value>
    </entry>
  </featureParams>
</trainingFeature>

```

- **KeyWords:** busca las palabras que especificó como pertenecientes a una lista personalizada, como *DaysOfWeek* o *Month*. De forma opcional, también especifica si el cruce debe distinguir mayúsculas de minúsculas; cuando se utiliza, el valor predeterminado es "verdadero".

```

<trainingFeature>
  <featureName>KeyWords</featureName>
  <featureParams>
    <entry>

```

```

    <key>KeyWordList</key>
    <value>Monday, Tuesday</value>
  </entry>
</entry>
  <key>CaseSensitive</key>
  <value>False</value>
</entry>
</featureParams>
</trainingFeature>

```

- **Substring:** extrae parte de una cadena como se especifica en los parámetros. También puede utilizarse para las extracciones de prefijo y sufijos.
  - **StartLocation:** izquierda o derecha. La posición en la que debe extraerse la subcadena. El valor predeterminado es Izquierda.
  - **StartPosition:** la posición de inicio de la subcadena. El valor predeterminado es 0.
  - **EndPosition:** la posición final para la subcadena. El valor predeterminado es 3.
  - **MinLength:** la longitud mínima de la palabra para la cual debe aplicarse esta función. El valor predeterminado es 3.

```

<trainingFeature>
  <featureName>Substring</featureName>
  <featureParams>
    <entry>
      <key>StartLocation</key>
    </entry>
    <entry>
      <key>StartPosition</key>
      <value>1</value>
    </entry>
    <entry>
      <key>EndPosition</key>
      <value>4</value>
    </entry>
    <entry>
      <key>MinLength</key>
    </entry>
  </featureParams>
</trainingFeature>

```

- **Características léxicas:** para especificar las propiedades de lexema
  - **FeatureWindow:** especifica la ventana para la generación de características

```

<trainingFeature>
  <featureName>FeatureWindow</featureName>
  <!-- Number of preceding tokens used to create the feature set.
  Default is 3 -->
  <entry>
    <key>Before</key>
    <value>1</value>
  </entry>
</trainingFeature>

```

```

    </entry>
    <!-- Number of succeeding tokens used to create the feature set.
    Default is 3 -->
    <entry>
      <key>After</key>
      <value>2</value>
    </entry>
  </trainingFeature>

```

A continuación encontrará un archivo completo de opciones de capacitación de ejemplo para entidades personalizadas:

```

<trainingOptions>
  <modelName>CustomModel</modelName>
  <modelType>CustomEntity</modelType>
  <modelDescription>CustomDiagnosesModel</modelDescription>

  <inputFilePath>C:/SpectrumIE/custom_model/Custom_Input.csv</inputFilePath>

  <testFilePath>C:/SpectrumIE/custom_model/Custom_Test.txt</testFilePath>

  <magicWord>DIAGNOSIS</magicWord>
  <language>English</language>

  <trainingFeatures>

  <!-- Lexical features-->
  <trainingFeature>
    <featureName>FeatureWindow</featureName>
    <featureParams>
      <entry>
        <key>Before</key>
        <value>1</value>
      </entry>
      <entry>
        <key>After</key>
        <value>2</value>
      </entry>
    </featureParams>
  </trainingFeature>

  <!-- Orthographic features-->
  <trainingFeature>
    <featureName>CaseIdentifier</featureName>
  </trainingFeature>

  <trainingFeature>
    <featureName>NumericIdentifier</featureName>
  </trainingFeature>
</trainingFeatures>
</trainingOptions>

```

## Capacitación del modelo de entidad personalizada

Después de crear un archivo de opciones, debe capacitar su modelo para identificar entidades personalizadas. Spectrum™ Technology Platform lo realiza mediante el comando CLI **iemodel train** en la página 37. Se utiliza un modelo capacitado para recuperar las entidades personalizadas.

## Evaluación del modelo de entidades personalizadas

Puede probar su modelo después de la capacitación para asegurarse de que el archivo de opciones de capacitación sea correcto y las entidades se extraigan como se espera. Para probar su modelo, use el comando de CLI **iemodel trainAndevaluate model** en la página 37.

## Extracción de las entidades personalizadas

La entidad personalizada capacitada, la cual ahora se encuentra disponible en la lista de entidades de la etapa **Entity Extractor**, se puede usar para extraer información relevante a partir de sus datos no estructurados.

Para obtener información sobre los pasos para extraer entidades preexistentes, consulte **Extracción de entidades preexistentes** en la página 16.

# 4 - Extracción de relaciones

## In this section

---

Introducción	24
Tipos de relación	25

# Introducción

*Relationship Extraction* le permite identificar los tipos de relación entre un par de entidades identificado en el contenido de origen. Analiza el contenido en lenguaje natural en el documento de origen e identifica los tipos de relación existentes entre todos los pares de entidad identificados.

## *Tipos de entidad admitidos*

Actualmente, los tipos de entidad admitidos para la extracción de relaciones son los siguientes:

- *Person*
- *Organization*
- *Location*

## Tipos de relación

RelationshipType	Tipo de entidad 1	Tipo de entidad 2	Relaciones cubiertas
<i>AffiliatedWith</i>	<i>Person</i>	<i>Organization</i>	<p>Indica cualquier relación profesional o académica entre las entidades de <i>Person</i> (Persona) y <i>Organization</i> (Organización).</p> <p>La relación puede ser cualquiera de las siguientes u otra similar:</p> <ul style="list-style-type: none"> <li>• <i>Person</i> está estudiando o estudió en <i>Organization</i></li> <li>• <i>Person</i> está trabajando o trabajó en <i>Organization</i></li> <li>• <i>Person</i> recibió una oferta para trabajar en <i>Organization</i></li> </ul> <p><b>Nota:</b> Esta es una lista indicativa de las relaciones que cubre este tipo.</p> <p>Por ejemplo:</p> <p>James has studied from the University of Toronto and works at ABC Corp.</p> <p>Aquí se pueden analizar dos relaciones:</p> <p>Entity1 = James, RelationshipType = AffiliatedWith, Entity2 = University of Toronto</p> <p>Entity1 = James, RelationshipType = AffiliatedWith, Entity2 = ABC Corp</p>

RelationshipType	Tipo de entidad 1	Tipo de entidad 2	Relaciones cubiertas
<i>LivesIn</i>	<i>Person</i>	<i>Location</i>	<p>Indica una relación entre las entidades de <i>Person</i> (Persona) y <i>Location</i> (Ubicación).</p> <p>La relación puede ser cualquiera de las siguientes:</p> <ul style="list-style-type: none"> <li>• <i>Person</i> se queda o se quedó en <i>Location</i></li> <li>• <i>Person</i> cambió a <i>Location</i></li> <li>• <i>Person</i> nació el <i>Location</i></li> <li>• <i>Person</i> falleció en <i>Location</i></li> </ul> <p><b>Nota:</b> Esta es una lista indicativa de las relaciones que cubre este tipo.</p> <p>Por ejemplo:</p> <p>John Jamison, a National Weather Service meteorologist in Galveston, reported that a massive hurricane was about to hit the East Coast the next day.</p> <p>Entity1 =John Jamison, RelationshipType = <i>LivesIn</i>, Entity2 = Galveston</p>
<i>OrgBasedIn</i>	<i>Organization</i>	<i>Location</i>	<p>Indica que la <i>Organization</i> (organización) tiene al menos una de sus oficinas en la <i>Location</i> (ubicación).</p> <p>La <i>Location</i> (ubicación) puede ser una sucursal, oficina de desarrollo, oficina central y similares.</p> <p>Por ejemplo:</p> <p>HSBC Holdings Plc. is headquartered in London, United Kingdom.</p> <p>Aquí se pueden analizar dos relaciones:</p> <p>Entity1 = HSBC Holdings Plc., RelationshipType = <i>OrgBasedIn</i>, Entity2 =London</p> <p>Entity1 = HSBC Holdings Plc., RelationshipType = <i>OrgBasedIn</i>, Entity2 = United States of America</p>

RelationshipType	Tipo de entidad 1	Tipo de entidad 2	Relaciones cubiertas
<i>LocatedIn</i>	<i>Location</i>	<i>Location</i>	<p>Indica la relación entre dos ubicaciones diferentes, donde una de las entidades se encuentra contenida geográficamente dentro de otra entidad.</p> <p><b>Ejemplo 1</b> Canberra is the capital of Australia.</p> <p>Aquí, Entity1 = Canberra, RelationshipType =<i>LocatedIn</i>, Entity2 =Australia</p> <p><b>Ejemplo 2</b> India has as its capital New Delhi.</p> <p>Aquí, Entity1 = India, RelationshipType =<i>LocatedIn</i>, Entity2 = New Delhi</p>
<i>Negative</i>	<i>Person</i> <i>Organization</i> <i>Location</i>	<i>Organization</i> <i>Location</i>	<p>Indica que ninguno de los tipos de relación mencionados anteriormente pudo analizarse entre las dos entidades correspondientes.</p> <p>Por ejemplo:</p> <p>New Delhi and New York are good places to live in.</p> <p>Al analizar este texto de entrada, no se analiza ninguno de los tipos de relación admitidos entre cualquier par de entidades identificadas. Por lo tanto, este se puede desglosar en tipos de relación <i>Negative</i> (negativos) entre las entidades identificadas:</p> <p>Entity1 = New Delhi, RelationshipType =<i>Negative</i>, Entity2 = New York</p>

**Nota:** Puede conectar una etapa **Splitter** en la salida de la etapa **Relationship Extractor** para extraer los tipos de relación identificados y el par de entidades correspondiente unidas por la relación. La etapa Splitter convierte los datos de salida jerárquicos de esta etapa en datos de salida planos.

### Ejemplo

En caso de un texto de entrada complejo, pueden analizarse múltiples combinaciones de tipo de relación posibles para la misma oración.

Por ejemplo:

James McCarthy has settled in New York, United States as director of ABC Technologies.

Cuando la etapa **Relationship Extractor** analiza este texto de entrada mediante el uso de los tipos de relación seleccionados en las opciones de la etapa, las relaciones encontradas son las siguientes:

- Relación 1** Entity1 = James McCarthy, Entity1 Type = *Person*, RelationshipType = *LivesIn*, Entity2 = New York, Entity2 Type = *Location*
- Relación 2** Entity1 = James McCarthy, Entity1 Type = *Person*, RelationshipType = *AffiliatedWith*, Entity2 = ABC Technologies, Entity2 Type = *Organization*
- Relación 3** Entity1 = ABC Technologies, Entity1 Type = *Organization*, RelationshipType = *OrgBasedIn*, Entity2 = United States, Entity2 Type = *Location*
- Relación 4** Entity1 = ABC Technologies, Entity1 Type = *Organization*, RelationshipType = *OrgBasedIn*, Entity2 = New York, Entity2 Type = *Location*
- Relación 5** Entity1 = James McCarthy, Entity1 Type = *Person*, RelationshipType = *LivesIn*, Entity2 = United States, Entity2 Type = *Location*
- Relación 6** Entity1 = New York, Entity1 Type = *Location*, RelationshipType = *LocatedIn*, Entity2 = United States, Entity2 Type = *Location*

# 5 - Comandos de la Utilidad de administración

## In this section

---

Comandos de la Utilidad de administración	30
iemodel delete	30
iemodel evaluate model	30
iemodel evaluate train_model	33
iemodel export	35
iemodel import	35
iemodel list	36
iemodel train	37
iemodel trainAndevaluate model	37

## Comandos de la Utilidad de administración

Los comandos de la utilidad de administración se utilizan para administrar, evaluar y capacitar los modelos del módulo Information Extraction.

### iemodel delete

El comando `iemodel delete` devuelve una lista de todos los modelos del módulo Information Extraction.

#### Uso

```
iemodel delete --n modelName
```

Requerido	Argumento	Descripción
Sí	--n <i>modelName</i>	Especifica el nombre del modelo que desea eliminar. Las rutas de acceso del directorio que especifique aquí es en relación con la ubicación donde ejecuta la Utilidad de administración.

#### Ejemplo

Este ejemplo elimina el modelo llamado "MyModel".

```
iemodel delete --n MyModel
```

### iemodel evaluate model

El comando `iemodel evaluate` evalúa un modelo del módulo Information Extraction que se ha capacitado previamente.

#### Uso

```
iemodel evaluate  
model --n modelName --t testFileName --o outputFileName --c categoryCount --d trueOrfalse
```

Requerido	Argumento	Descripción
Sí	<code>--n modelName</code>	Especifica el nombre y la ubicación del modelo que desea evaluar. Las rutas de acceso del directorio que especifique aquí es en relación con la ubicación donde ejecuta la Utilidad de administración.
Sí	<code>--t testFileName</code>	Especifica el nombre y la ubicación del archivo de prueba utilizado para evaluar el modelo.
No	<code>--o outputFileName</code>	Especifica el nombre y la ubicación del archivo de salida que guardará los resultados de la evaluación.
No	<code>--c categoryCount</code>	Especifica el número de categorías en el modelo; debe ser un valor numérico.

**Nota:** Válido solo para el modelo de clasificación de texto.

No	<code>--d trueOrfalse</code>	<p>Especifica si se visualiza una tabla con análisis detallado con relación a la entidad; el valor debe ser <code>true</code> o <code>false</code>, como se muestra a continuación:</p> <p><b>true</b></p> <p style="padding-left: 40px;">Se requieren resultados detallados de la evaluación.</p> <p><b>false</b></p> <p style="padding-left: 40px;">No se requieren resultados detallados de la evaluación.</p> <p>El valor predeterminado es <code>false</code>.</p> <p>En la tabla <i>Resultados de la evaluación del modelo</i>, y en la <i>Matriz de confusión</i> con sus columnas, según se describe a continuación, aparecen los recuentos por entidad.</p> <p><b>Nota:</b> Si el comando se ejecuta sin este argumento o con el valor de argumento <code>falso</code>, la tabla <i>Resultados de la evaluación del modelo</i> y la <i>Matriz de confusión</i> no se muestran. Solo aparecen las <i>Estadísticas de evaluación de modelo</i>.</p>
----	------------------------------	--

## Salida

### Estadísticas de evaluación de modelo

La ejecución de este comando muestra las siguientes estadísticas de la evaluación en formato tabular:

- **Precisión:** es una medida de la exactitud. Precision define la proporción de tuplas identificadas correctamente.
- **Recuperación:** es una medida de integridad de los resultados. La recuperación se puede definir como una fracción de instancias relevantes que se recupera.

- **Medida F1:** es la medida de la exactitud de una prueba. El cálculo del puntaje de F1 considera la precisión y la recuperación de la prueba. Se puede interpretar como el promedio ponderado de precisión y recuperación, donde el puntaje de F1 alcanza el mejor valor en 1 y el peor en 0.
- **Exactitud:** mide el nivel de corrección de los resultados. Define la cercanía del valor medido al valor conocido.

### Resultados de evaluación de modelo

Si el comando se ejecuta con el argumento `--d true`, los recuentos de cruce de todas las entidades aparecen en formato tabular. Las columnas en la tabla son:

<b>Recuento de entrada</b>	La cantidad de ocurrencias de la entidad en los datos de entrada.
<b>Recuento de coincidencia incorrecta</b>	La cantidad de veces que hubo error en el cruce de la entidad.
<b>Recuento de cruces</b>	La cantidad de veces que el cruce de la entidad se hizo correctamente.

### Matriz de confusión

La *Matriz de confusión* (mostrada abajo) permite visualizar el rendimiento de un algoritmo. En esta se ilustra el rendimiento de un modelo de clasificación.



La columna representa las instancias en una clase prevista, en tanto que la fila representa las instancias en una clase real. Estos son algunos términos relacionados con la matriz de confusión:

<b>Real</b>	La cantidad de ocurrencias de la entidad en la clase real.
<b>Prevista</b>	La cantidad de ocurrencias de la entidad en la clase prevista.
<b>TP</b>	<b>Verdadero positivo:</b> La cantidad de ocurrencias de entidad predichas como positivas y que resultaron ser realmente verdaderas.
<b>TN</b>	<b>Verdadero negativo:</b> La cantidad de ocurrencias de entidad predichas como negativas pero que resultaron ser verdaderas.
<b>FP</b>	<b>Falso positivo:</b> La cantidad de ocurrencias de entidad predichas como positivas pero que resultaron ser falsas.
<b>FN</b>	<b>Falso negativo:</b> La cantidad de ocurrencias de entidad predichas como negativas y que resultaron ser realmente falsas.

#### Ejemplo

Este ejemplo:

- Evalúa el modelo llamado "MyModel"
- Utiliza un archivo de prueba llamado "ModelTestFile" en la misma ubicación
- Guarda el resultado de la evaluación en un archivo llamado "MyModelTestOutput"
- Especifica un recuento de categoría de 4
- Especifica que se requiere un análisis detallado de la evaluación.

```
iemodel evaluate model --n MyModel --t
C:\Spectrum\IEModels\ModelTestFile --o
C:\Spectrum\IEModels\MyModelTestOutput --c 4 --d true
```

## iemodel evaluate train\_model

El comando `iemodel evaluate train_model` evalúa y capacita un modelo existente del **Módulo Information Extraction**. Esta función no se puede realizar en un modelo nuevo.

**Nota:** Para obtener resultados óptimos en la evaluación y la capacitación de un **Módulo Information Extraction** existente, use este comando: `iemodel trainAndevaluate model`. Para obtener detalles, consulte [iemodel trainAndevaluate model](#) en la página 37.

### Uso

```
iemodel evaluate
train_model --f trainingOptionsFile --u trueOrFalse --o outputFileName --c categoryCount --d trueOrfalse
```

Requerido	Argumento	Descripción
Sí	--f <i>trainingOptionsFile</i>	Especifica el nombre y la ubicación del archivo con las opciones de capacitación que se utilizan para capacitar el modelo. Las rutas de acceso del directorio que especifique aquí es en relación con la ubicación donde ejecuta la Utilidad de administración.
No	--u <i>overWritelfExists</i>	Especifica si se debe sobrescribir el modelo capacitado existente (en caso de que exista uno). <i>TrueOrFalse</i> es una de las siguientes opciones: <ul style="list-style-type: none"> <li><b>verdadero</b>                      Sobrescribe el modelo existente.</li> <li><b>falso</b>                              No sobrescribe el modelo existente.</li> </ul>
No	--o <i>outputFileName</i>	Especifica el nombre y la ubicación del archivo de salida que guardará los resultados de la evaluación.
No	--c <i>categoryCount</i>	Especifica el número de categorías en el modelo; debe ser un valor numérico. <p style="text-align: right;"><b>Nota:</b> Válido solo para el modelo de clasificación de texto.</p>
No	--d <i>trueOrfalse</i>	Especifica si se visualiza una tabla con análisis detallado con relación a la entidad; el valor debe ser <code>true</code> o <code>false</code> , como se muestra a continuación: <p><b>true</b></p>

Requerido	Argumento	Descripción
		Se requieren resultados detallados de la evaluación.
	<b>false</b>	No se requieren resultados detallados de la evaluación.
		El valor predeterminado es <code>false</code> .
		La tabla <i>Resultados de la evaluación del modelo</i> , con las columnas que se describen más abajo, muestra los recuentos por entidad.
		<b>Nota:</b> Si el comando se ejecuta sin este argumento o con el valor de argumento <code>false</code> , la tabla Resultados de la evaluación del modelo no se muestra. Solo aparecen las estadísticas de evaluación de modelo.

### Salida

#### Estadísticas de evaluación de modelo

La ejecución de este comando muestra las siguientes estadísticas de la evaluación en formato tabular:

- Precisión
- Recuperación
- Medida F1

#### Resultados de evaluación de modelo

Si el comando se ejecuta con el argumento `--d true`, los recuentos de cruce de todas las entidades aparecen en formato tabular. Las columnas en la tabla son:

<b>Recuento de entrada</b>	La cantidad de ocurrencias de la entidad en los datos de entrada.
<b>Recuento de coincidencia incorrecta</b>	La cantidad de veces que hubo error en el cruce de la entidad.
<b>Recuento de cruces</b>	La cantidad de veces que el cruce de la entidad se hizo correctamente.

#### Ejemplo

Este ejemplo:

- Utiliza un archivo de opciones de capacitación llamado "ModelTrainingFile" que se encuentra en "C:\Spectrum\IEModels"
- Sobrescribe todo archivo de salida existente que tenga el mismo nombre
- Guarda el resultado de la evaluación en un archivo llamado "MyModelTestOutput"
- Especifica un recuento de categoría de 4
- Especifica que se requiere un análisis detallado de la evaluación.

```
iemodel evaluate train_model --f
C:\Spectrum\IEModels\ModelTrainingFile --u true --o
C:\Spectrum\IEModels\MyModelTestOutput --c 4 --d true
```

## iemodel export

El comando `iemodel export` exporta un modelo del módulo Information Extraction y sus metadatos.

### Uso

```
iemodel export --n modelName --o outputDirectory
```

Requerido	Argumento	Descripción
Sí	--n <i>modelName</i>	Especifica el nombre del modelo que desea exportar. Las rutas de acceso del directorio que especifique aquí es en relación con la ubicación donde ejecuta la Utilidad de administración.
Sí	--o <i>outputDirectory</i>	Especifica la ubicación de la carpeta que guardará el modelo exportado y sus metadatos.

### Ejemplo

Este ejemplo exporta un modelo denominado `MyModel` que coloca el resultado en una carpeta llamada "MyModelExport", ubicada en "C:\Spectrum\IEModels\MyModelExport".

```
iemodel export --n MyModel --o
C:\Spectrum\IEModels\MyModelExport
```

## iemodel import

El comando `iemodel import` importa un modelo del módulo Information Extraction y sus metadatos.

### Uso

```
iemodel import --n modelName --o inputDirectory --u trueOrFalse
```

Requerido	Argumento	Descripción				
Sí	<code>--n modelName</code>	Especifica el nombre del modelo que desea importar. Las rutas de acceso del directorio que especifique aquí es en relación con la ubicación donde ejecuta la Utilidad de administración.				
Sí	<code>--o inputDirectory</code>	Especifica la ubicación de la carpeta que guardará el modelo importado y sus metadatos.				
No	<code>--u overWritelfExists</code>	Especifica si se debe sobrescribir el modelo existente (en caso de que exista uno). <i>TrueOrFalse</i> es una de las siguientes opciones: <table border="0" style="margin-left: 20px;"> <tr> <td><b>verdadero</b></td> <td>Sobrescribe el modelo existente.</td> </tr> <tr> <td><b>falso</b></td> <td>No sobrescribe el modelo existente.</td> </tr> </table>	<b>verdadero</b>	Sobrescribe el modelo existente.	<b>falso</b>	No sobrescribe el modelo existente.
<b>verdadero</b>	Sobrescribe el modelo existente.					
<b>falso</b>	No sobrescribe el modelo existente.					

**Ejemplo**

Este ejemplo importa un modelo denominado `MyModel` que guarda el modelo en una carpeta llamada "MyModelExport", ubicada en "C:\Spectrum\IEModels\MyModelExport". Además sobrescribe todo modelo existente que tenga el mismo nombre.

```
iemodel import --n MyModel --o
C:\Spectrum\IEModels\MyModelExport --u true
```

## iemodel list

El comando `iemodel list` devuelve una lista de todos los modelos del módulo Information Extraction.

**Uso**

```
iemodel list
```

**Ejemplo**

Este ejemplo enumera todos los modelos.

```
iemodel list
```

## iemodel train

El comando `iemodel train` capacita un modelo de módulo Information Extraction. Este invoca su archivo de opciones de capacitación, que apunta al archivo de entrada y aplica las opciones que usted definió.

### Uso

```
iemodel train --f trainingOptionsFile --u trueOrFalse
```

Requerido	Argumento	Descripción
Sí	<code>--f <i>trainingOptionsFile</i></code>	Especifica el nombre y la ubicación del archivo con las opciones de capacitación que se utilizan para capacitar el modelo. Las rutas de acceso del directorio que especifique aquí es en relación con la ubicación donde ejecuta la Utilidad de administración.
No	<code>--u <i>trueOrFalse</i></code>	Especifica si se debe sobrescribir el modelo existente, si ya existe uno con el mismo nombre en el servidor, donde <i>TrueOrFalse</i> sea una de las siguientes opciones: <b>true</b> Sobrescribe el modelo existente. <b>false</b> No sobrescribe el modelo existente.

### Ejemplo

Este ejemplo enseña un modelo incluido en el archivo *TrainingOptions.xml*, el cual se almacena en la unidad C: y sobrescribe todo modelo existente con el mismo nombre.

```
iemodel train --f c:/TrainingOptions.xml --u true
```

## iemodel trainAndevaluate model

El comando `iemodel trainAndevaluate model` evalúa y capacita nuevos modelos y modelos existentes. En el caso de los modelos existentes, debe sobrescribirlos con los modelos recién capacitados, y para hacerlo debe usar "true" para el argumento `--u` del comando.

Este comando invoca su archivo de opciones de capacitación y ofrece un archivo de salida opcional con resultados de evaluaciones, en caso de que opte por producir ese archivo.

### Uso

```
ie model trainAndevaluate
```

```
model --f trainingOptionsFile --u trueOrFalse --o outputFileName --c categoryCount --d trueOrfalse
```

Requerido	Argumento	Descripción
Sí	--f <i>trainingOptionsFile</i>	Especifica el nombre y la ubicación del archivo con las opciones de capacitación que se utilizan para capacitar el modelo. Las rutas de acceso del directorio que especifique aquí es en relación con la ubicación donde ejecuta la Utilidad de administración.
No	--u <i>overWriteIfExists</i>	Especifica si se debe sobrescribir el modelo capacitado existente (en caso de que exista uno). <b>verdadero</b> Sobrescribe el modelo existente. <b>falso</b> No sobrescribe el modelo existente.
No	--o <i>outputFileName</i>	Especifica el nombre y la ubicación del archivo de salida que guardará los resultados de la evaluación.
No	--c <i>categoryCount</i>	Especifica el número de categorías en el modelo; debe ser un valor numérico.  <b>Nota:</b> Válido solo para el modelo de clasificación de texto.
No	--d <i>trueOrfalse</i>	Especifica si se visualiza una tabla con análisis detallado con relación a la entidad; el valor debe ser <code>true</code> o <code>false</code> , como se muestra a continuación: <b>true</b>  Se requieren resultados detallados de la evaluación. <b>false</b>  No se requieren resultados detallados de la evaluación.  El valor predeterminado es <code>false</code> .  En la tabla <i>Resultados de la evaluación del modelo</i> , y en la <i>Matriz de confusión</i> con sus columnas, según se describe a continuación, aparecen los recuentos por entidad.  <b>Nota:</b> Si el comando se ejecuta sin este argumento o con el valor de argumento <code>falso</code> , la tabla <i>Resultados de la evaluación del modelo</i> y la <i>Matriz de confusión</i> no se muestran. Solo aparecen las <i>Estadísticas de evaluación de modelo</i> .

## Salida

### Estadísticas de evaluación de modelo

La ejecución de este comando muestra las siguientes estadísticas de la evaluación en formato tabular:

- **Precisión:** es una medida de la exactitud. Precision define la proporción de tuplas identificadas correctamente.
- **Recuperación:** es una medida de integridad de los resultados. La recuperación se puede definir como una fracción de instancias relevantes que se recupera.
- **Medida F1:** es la medida de la exactitud de una prueba. El cálculo del puntaje de F1 considera la precisión y la recuperación de la prueba. Se puede interpretar como el promedio ponderado de precisión y recuperación, donde el puntaje de F1 alcanza el mejor valor en 1 y el peor en 0.
- **Exactitud:** mide el nivel de corrección de los resultados. Define la cercanía del valor medido al valor conocido.

### Resultados de evaluación de modelo

Si el comando se ejecuta con el argumento `--d true`, los recuentos de cruce de todas las entidades aparecen en formato tabular. Las columnas en la tabla son:

<b>Recuento de entrada</b>	La cantidad de ocurrencias de la entidad en los datos de entrada.
<b>Recuento de coincidencia incorrecta</b>	La cantidad de veces que hubo error en el cruce de la entidad.
<b>Recuento de cruces</b>	La cantidad de veces que el cruce de la entidad se hizo correctamente.

### Matriz de confusión

La *Matriz de confusión* (mostrada abajo) permite visualizar el rendimiento de un algoritmo. En esta se ilustra el rendimiento de un modelo de clasificación.



La columna representa las instancias en una clase prevista, en tanto que la fila representa las instancias en una clase real. Estos son algunos términos relacionados con la matriz de confusión:

<b>Real</b>	La cantidad de ocurrencias de la entidad en la clase real.
<b>Prevista</b>	La cantidad de ocurrencias de la entidad en la clase prevista.
<b>TP</b>	<b>Verdadero positivo:</b> La cantidad de ocurrencias de entidad predichas como positivas y que resultaron ser realmente verdaderas.
<b>TN</b>	<b>Verdadero negativo:</b> La cantidad de ocurrencias de entidad predichas como negativas pero que resultaron ser verdaderas.
<b>FP</b>	<b>Falso positivo:</b> La cantidad de ocurrencias de entidad predichas como positivas pero que resultaron ser falsas.
<b>FN</b>	<b>Falso negativo:</b> La cantidad de ocurrencias de entidad predichas como negativas y que resultaron ser realmente falsas.

**Ejemplo**

Este ejemplo:

- Utiliza un archivo de opciones de capacitación llamado "ModelTrainingFile" que se encuentra en "C:\Spectrum\IEModels"
- Sobrescribe todo archivo de salida existente que tenga el mismo nombre
- Guarda el resultado de la evaluación en un archivo llamado "MyModelTestOutput"
- Especifica un recuento de categoría de 4
- Especifica que se requiere un análisis detallado de la evaluación.

```
iemodel trainAndevaluate model --f  
C:\Spectrum\IEModels\ModelTrainingFile --u true --o  
C:\Spectrum\IEModels\MyModelTestOutput --c 4 --d true
```

# 6 - Referencia de etapas

## In this section

---

Componentes de Information Extraction	42
Read from Documents	42
Entity Extractor	47
Text Categorizer	50
Extractor de relaciones	52

## Componentes de Information Extraction

El módulo Information Extraction incluye las siguientes etapas.

- **Read From Documents:** lee datos de entrada sin estructura desde varios formatos de archivo y extrae el contenido.
- **Entity Extractor:** extrae entidades tales como nombres y direcciones desde datos sin estructura pasados como cadenas de caracteres.
- **Text Categorizer:** asigna categorías personalizadas al contenido no estructurado o texto sin formato (como correos electrónicos, artículos noticiosos y comentarios) según cuánto de dicho contenido tiene el material para esa categoría.
- **Relationship Extractor:** extrae relaciones entre entidades.

## Read from Documents

Read from Documents es una etapa de origen que lee datos de entrada sin estructura desde varios formatos de archivo y extrae el contenido. Las fuentes posibles incluyen documentos legales, comentarios del usuario, revisiones de productos, artículos noticiosos, blogs, redes sociales, etc. Read from Documents también extrae campos de metadatos, como la fecha de creación y el autor. Después de extraer los datos, estos se pueden utilizar para varios tipos de procesamientos, lo que incluye extracción de entidades y manipulación de cadenas, entre otras. Los datos también se pueden usar para construir índices de búsqueda para buscar texto sin estructura.

**Nota:** Cada documento se considera como un registro para esta etapa.

## Input

Los datos de entrada de Read from Documents son un archivo o una carpeta. Esta etapa admite los siguientes tipos de archivo:

- Texto
- PDF
- Microsoft Outlook
- Microsoft Word
- HTML

Read from Documents realiza tres tipos de extracciones:

- Documento: use el documento completo
- Página: use una página específica de un documento
- Selectiva: use una parte seleccionada de un documento
- Marcadores: use los marcadores de un documento PDF

Read from Documents forma parte del módulo Information Extraction.

## Opciones

### *Ficha Propiedades del archivo*

La tabla a continuación muestra las opciones que controlan el tipo de información devuelta por Read from Documents.

**Tabla 1: Opciones de ReadfromDocuments**

Opción	Descripción
Server name (Nombre de servidor)	Especifica el nombre del servidor de Spectrum Technology Platform que se utiliza.
Nombre de carpeta o archivo	La ruta de acceso y el nombre de la carpeta o el documento de origen. Si desea apuntar a una carpeta, use un asterisco como carácter comodín ("*"), para seleccionar los archivos en la carpeta. Si desea apuntar a varios archivos del mismo tipo dentro de una carpeta, use el carácter comodín más la e de archivo ("*.pdf").
Tipo de archivo	El tipo de archivo del documento de origen, que se seleccionará automáticamente después de elegir una fuente: <ul style="list-style-type: none"> <li>• Texto</li> <li>• PDF</li> <li>• Microsoft Outlook</li> <li>• Microsoft Word</li> <li>• HTML</li> </ul>

Opción	Descripción
Tipo de extracción	<p><b>Documentación</b> Use el documento completo.</p> <p><b>Página</b> Use una página específica de un documento.</p> <p><b>Selección</b> Use una parte seleccionada de un documento.</p> <p><b>Marcadores</b> Use los marcadores de un documento PDF.</p>
Selección de página	Solo con el tipo de extracción <b>Página</b> . Seleccione todas las páginas o un rango de las mismas.
Extracción seleccionada	Solo con el tipo de extracción <b>Selección</b> . Especifica el tipo de búsqueda.
Especifique texto	Solo con el tipo de extracción <b>Selección</b> . Especifica el texto que se va a buscar.
Excluir texto de inicio	Solo con el tipo de extracción <b>Selección</b> y la opción <b>Texto inicial</b> . Omite la cadena ingresada desde los datos devueltos.
Especificar texto final	Solo con el tipo de extracción <b>Selección</b> . Especifica el texto final que se va a buscar.
Excluir texto final	Solo con el tipo de extracción <b>Selección</b> . Omite la cadena ingresada desde el final de los datos devueltos.
Arrojar selección	Solo con el tipo de extracción <b>Selección</b> . Especifica cuántos párrafos se deben devolver para cada resultado. Por ejemplo, si elige "2", los datos devueltos para cada resultado incluirán el párrafo donde está el resultado más el párrafo posterior, totalizando dos párrafos. El valor predeterminado es 1. No es válido cuando se especifica el texto final.

### *Ficha Campos*

Haga clic en **Regenerar** para definir los campos de entrada.

**Tabla 2: Opciones de datos de salida**

Opción	Descripción
Nombre de atributo	Muestra el atributo que más se parece al campo de entrada. Por ejemplo, si uno de sus campos contiene información de fecha y lo llama "Fecha", podrá ver el atributo "Fecha" asignado a dicho campo. Esta columna no se puede editar.
Nombre	El nombre del campo. Esta columna se puede editar.
Tipo	El tipo de datos del campo.
Incluir	Especifica qué campos se van a incluir en un índice de búsqueda.

## Salida

La etapa Read from Documents posee dos puertos de salida. Un puerto captura los datos leídos por la etapa y devueltos a partir de los criterios ingresados. Estos datos pueden incluir texto sin formato o metadatos (como autor, idioma, fecha de creación, etc.). Este puerto se puede conectar a cualquier etapa que lea datos entrantes, como por ejemplo, Write to File o Write to XML, así como las etapas principales Validate Address o Write to Search Index. También se puede conectar a la etapa Information Extractor, si desea que se devuelva información acerca de ciertos tipos de entidades presentes en el documento. Cuando selecciona el tipo de extracción Documento los resultados incluirán datos planos; cuando selecciona el tipo de extracción Página o Selección, los resultados incluirán datos jerárquicos.

El otro puerto recopila todos los recursos que el flujo de datos no procesó correctamente. Este se denomina Puerto de error, y los registros que pasan por este puerto hacia el receptor se consideran incorrectos. Capturar registros incorrectos le puede ayudar a identificar el problema con aquellos registros. Cuando adjunta un receptor al puerto de error, el archivo de salida que se origina contendrá todos los campos de los registros malformados. También incluirá el campo Motivo que especifica el motivo por el que falló el registro.

**Tabla 3: Resultados de Unstructured Reader**

Nombre de campo	Descripción / Valores válidos
Autor	Normalmente contiene el nombre de la persona que creó o actualizó el documento. Esta información forma parte de los metadatos del documento.
Bookmark	Contiene todos los marcadores del archivo de entrada PDF. Solo para tipos de extracción de marcador.
BookmarkNo	Contiene todos los marcadores del archivo de entrada PDF. Solo para tipos de extracción de marcador.
ContentLength	Indica la longitud del documento. Este valor varía según el tipo de extracción seleccionada: <b>Documento</b> El número de páginas en el documento. <b>Página</b> "1", para representar la única página del documento.
Contenido	Varía según el tipo de extracción. Por ejemplo, los tipos de extracción de documento generarán el documento completo como datos planos. Los tipos de extracción de página, selección y marcadores generarán datos jerárquicos.
ContentType	Indica el tipo de documento que se leyó, por ejemplo, PDF, .txt, etc.
Creador	Normalmente contiene el nombre de la persona que creó el documento. Esta información forma parte de los metadatos del documento.
Fecha	Indica la fecha de creación o última actualización del documento.
Palabras clave	Contiene todas las palabras clave proporcionadas en los metadatos del documento.
Idioma	Indica el idioma en que se escribió el documento.
NPages	Indica el número de páginas en el documento.

Nombre de campo	Descripción / Valores válidos
PageContents	Incluye los contenidos de las páginas seleccionadas. Solo para tipos de extracción de página.
PageNo	Contiene el número de página para el marcador. Solo para tipos de extracción de página.
Elemento principal	Contiene la ruta del marcador, similar a XPath de un archivo XML. Solo para tipos de extracción de marcador.
ResourceName	Indica el nombre de archivo del documento.
SectionContents	Incluye los contenidos de la sección seleccionada. Solo para tipos de extracción de selección.
SectionNo	Indica el número de la sección dentro del documento. Solo para tipos de extracción de selección.
Asunto	Contiene el asunto del documento que se proporcionó en los metadatos del documento.
Título	Contiene el título del documento que se proporcionó en los metadatos del documento.

## Entity Extractor

Entity Extractor extrae entidades como los nombres y las direcciones de las cadenas de datos no estructurados (también denominados como texto sin formato).

Es posible que no se devuelvan todas las entidades para un tipo seleccionado, ya que la precisión varía según el tipo de documento de entrada. Dado que Entity Extractor utiliza procesamiento de lenguaje natural, una cadena de caracteres que contiene una oración gramaticalmente correcta de

un artículo noticioso o blog tendría una devolución de nombres más precisa que una simple lista de nombres y fechas.

## Parámetros d

**Entity Extractor** toma cadenas de datos no estructuradas como datos de entrada. También puede usar la etapa **Read from Documents** como datos de entrada si desea extraer entidades desde un documento sin estructura. La etapa **Read from Documents** permite leer el documento y devuelve texto según la configuración definida por el usuario. La etapa **Entity Extractor** extrae la información requerida desde el texto según las entidades seleccionadas.

**Tabla 4: Formato de entrada**

Nombre de campo	Descripción
PlainText	La cadena de datos no estructurada desde la cual desea extraer información.

## Opciones

Las opciones de Entity Extractor le permiten seleccionar las entidades a partir de las cuales desea extraer información de la cadena de caracteres de entrada. De manera predeterminada, puede extraer información si usa *Person* (Persona) y *Address* (Dirección) como los tipos de entidad. Sin embargo, puede utilizar la función **Agregado rápido** y seleccionar una de las 15 entidades preconfiguradas o todas estas.

Nombre de la opción	Descripción
Invaldar opciones predeterminadas del sistema con los siguientes valores	<p>Seleccione la casilla de verificación para anular los tipos de entidad predeterminados <i>Address</i> y <i>Person</i>.</p> <p>Cuando selecciona la casilla de verificación, el botón <b>Agregado rápido</b> se activa. Haga clic en este botón y seleccione las entidades que necesita para extraer el texto.</p> <p>Las entidades seleccionadas se añaden a la lista <b>Tipo de entidad</b>.</p>

Nombre de la opción	Descripción
Tipo de entidad	<p>Especifica el tipo de datos que desea extraer de la cadena sin estructura.</p> <p><b>Address</b></p> <p><b>CreditCard</b></p> <p><b>Date</b></p> <p><b>Email</b></p> <p><b>HashTag</b></p> <p><b>ISBN</b></p> <p><b>Location</b></p> <p><b>Mention</b></p> <p><b>Organization</b></p> <p><b>Person</b></p> <p><b>Phone</b></p> <p><b>ProperNouns</b></p> <p><b>SSN</b></p> <p><b>WebAddress</b></p> <p><b>ZipCode</b></p>
Recuento de entidades de salida	<p>Especifica si se debe devolver un recuento de la cantidad de veces que ocurrió una entidad determinada en los datos de salida.</p> <p><b>true</b>      Obtenga un recuento de las entidades encontradas en la cadena sin estructura.</p> <p><b>false</b>      No obtenga un recuento de las entidades encontradas en la cadena sin estructura.</p>

## Output

Los datos de salida de **Entity Extractor** son una lista de las entidades de comparación encontradas en la cadena de caracteres de entrada. Por ejemplo, si seleccionó un tipo de entidad como "Persona", el resultado contendría una lista de los nombres de personas encontrados en la cadena de caracteres de entrada. Igualmente, si seleccionó un **Tipo de entidad** como "Fecha", el resultado será una lista de las fechas encontradas en la cadena de caracteres de entrada.

Cada entidad, ya sea nombre, dirección o fecha, se devuelve solo una vez incluso si la entidad aparece varias veces en la cadena de caracteres de entrada.

Para ver la cantidad de veces que la entidad apareció en la cadena de caracteres de entrada puede seleccionar la opción **Conteo de entidades de salida** en la ventana **Opciones de Entity Extractor**.

Nombre de campo	Descripción
Text	El texto extraído desde la cadena.
Type	El tipo de entidad del texto extraído. Una de las siguientes: <b>Address</b> <b>CreditCard</b> <b>Date</b> <b>Email</b> <b>HashTag</b> <b>ISBN</b> <b>Location</b> <b>Mention</b> <b>Organization</b> <b>Person</b> <b>Phone</b> <b>ProperNouns</b> <b>SSN</b> <b>WebAddress</b> <b>ZipCode</b>
Count	Si la opción para devolver un recuento está activada, este campo contiene la cantidad de veces que aparece una entidad determinada en los datos de entrada. Por ejemplo, si elige devolver las entidades <code>Name</code> (Nombre) y el texto de entrada contiene cinco instancias del nombre <code>John</code> , dicho nombre se incluirá solo una vez en los datos de salida, con <code>Name</code> como el tipo de entidad y "5" como el recuento de resultados.

## Text Categorizer

Esta etapa lo ayuda a asignar categorías personalizadas a contenido no estructurado o texto sin formato (como correos electrónicos, artículos noticiosos y comentarios), según el nivel de comparación del contenido. La etapa incluye las categorías definidas, desde las cuales puede

seleccionar la que necesita para su categorización. Sin embargo, debe crear estas categorías mediante la capacitación de un modelo categorizador con sus datos. Para obtener más detalles, consulte [Introducción a la Categorización de texto](#) en la página 7.

## Parámetros d

La etapa toma cadenas de datos no estructuradas como entrada. También puede usar la etapa **Read from Documents** como datos de entrada si desea categorizar texto desde un documento sin estructura. La etapa **Read from Documents** permite leer el documento y devuelve texto según la configuración definida por el usuario. La lectura se realiza en la etapa **Text Categorizer** y le permite obtener los datos de salida que desea.

**Tabla 5: Formato de entrada**

Nombre de campo	Descripción
PlainText	La cadena de datos no estructurada desde la cual desea extraer información.

## Opciones

Las **Opciones de Text Categorizer** le permiten elegir los parámetros según los cuales desea clasificar su cadena de datos de entrada. Puede seleccionar el modelo para categorización y la cantidad de niveles de comparación en los que desea que se encuentren los datos de salida. Por ejemplo, solo el cruce más cercano o el que le sigue en cercanía.

Nombre de la opción	Descripción
Invaldar opciones predeterminadas del sistema con los siguientes valores	Para anular la opción predeterminada y seleccionar el categorizador a partir de la lista desplegable <b>Nombre del categorizador</b> .
Nombre del categorizador	Especifica el modelo que se debe usar para la categorización de texto. Detalla todos los modelos que capacitó en la fase de categorización de texto. <b>Nota:</b> Para obtener más información, consulte <a href="#">Capacitación del modelo</a> en la página 12.

Nombre de la opción	Descripción
Conteo de categorías	<p>El recuento de niveles de comparación de categoría que desea en los datos de salida. Por ejemplo, seleccione 1 para mostrar solo el cruce más cercano y 2 para mostrar el que le sigue en cercanía.</p> <p><b>Nota:</b> El valor máximo corresponde a la cantidad de clases diferentes especificadas durante la capacitación del modelo.</p>

## Output

Los datos de salida indican las categorías en las cuales el contenido de la cadena de caracteres de entrada se clasifica y el rango de dicha categoría. El rango indica el nivel de cercanía del cruce entre el contenido de entrada y la categoría. Por ejemplo, 1 significa que es el cruce más cercano con la categoría y 2 significa el cruce que le sigue en cercanía.

Nombre de campo	Descripción
Category	La categoría prevista para cada registro en el archivo de entrada.
Rank	El rango de categorías de la calificación más alta a la más baja.

## Extractor de relaciones

La etapa **Relationship Extractor** le permite identificar los tipos de relación entre las entidades identificadas en el contenido de origen.

La etapa **Relationship Extractor** identifica:

1. Entidad 1
2. Tipo de entidad 1
3. Tipo de relación
4. Entidad 2
5. Tipo de entidad 2

**Importante:** La etapa intenta alcanzar la máxima precisión posible mientras identifica los tipos de relación entre cualquiera de las dos entidades en el texto de entrada. Sin embargo, las relaciones imprecisas entre las dos entidades también se pueden identificar mientras se analizan oraciones complicadas en el texto de entrada.

## Parámetros d

La etapa **Relationship Extractor** toma cadenas de datos en lenguaje natural como datos de entrada e identifica las entidades y los tipos de relación existentes entre cada par de entidades.

Use la etapa **Read from Documents** como una etapa de origen si el texto de entrada proviene de un documento no estructurado. La etapa **Read from Documents** permite leer el documento y devuelve texto según la configuración definida por el usuario.

La etapa **Relationship Extractor** luego identifica todas las entidades y el tipo de relación existente entre cada par de entidades.

**Tabla 6: Formato de entrada**

Nombre de campo	Descripción
PlainText	La cadena de datos no estructurada a partir de la cual desea identificar los tipos de relación existentes entre cada par de entidades.

## Opciones

Las opciones de la etapa **Relationship Extractor** le permiten especificar qué tipos de relación desea identificar en el texto de entrada.

De manera predeterminada, los tipos de relación identificados son los siguientes:

1. *AffiliatedWith*
2. *LivesIn*
3. *OrgBasedIn*
4. *LocatedIn*

Nombre de la opción	Descripción
Invaldar opciones predeterminadas del sistema con los siguientes valores	<p>Seleccione la casilla de verificación para anular los tipos de relación predeterminados identificados y especifique qué tipos de relación desea identificar y extraer a partir del texto de entrada.</p> <p>Cuando selecciona la casilla de verificación, el botón <b>Agregado rápido</b> se activa. Haga clic en <b>Agregado rápido</b> para seleccionar los tipos de relación que desea identificar en el texto.</p> <p>Las entidades seleccionadas se añaden a la lista <b>Tipo de relación</b>.</p>

## Output

Los datos de salida de **Relationship Extractor** son una lista de conjuntos de relaciones identificadas entre pares de entidades encontradas en la cadena de caracteres de entrada.

Por ejemplo, si está en las opciones de la etapa y seleccionó extraer los tipos de relación *LivesIn* y *OrgBasedIn*, los datos de salida contienen una lista de todos los conjuntos de *Person LivesIn Location* y *Organization OrgBasedIn Location* identificados en el texto de entrada.

Cada par de entidad con su tipo de relación relacionado se incluye solo una vez.

Para cada conjunto de entidades extraído y su relación, la información extraída es:

Nombre de campo	Descripción
Entity1	La primera entidad de un par de entidades extraídas a partir del texto de entrada.
Entity1 Type	<p>El tipo de entidad de la primera entidad del par de entidades extraídas a partir del texto de entrada.</p> <p>El tipo de entidad es uno de los siguientes:</p> <ul style="list-style-type: none"> <li>• <i>Person</i></li> <li>• <i>Organization</i></li> <li>• <i>Location</i></li> </ul>

Nombre de campo	Descripción
Type	<p>El tipo de relación identificado entre Entidad 1 y Entidad 2.</p> <p>Para obtener más información sobre los tipos de relación, consulte <a href="#">Tipos de relación</a> en la página 25.</p> <p><b>Nota:</b> Solo se identifican e incluyen los tipos de relación seleccionados para extracción en las opciones de la etapa.</p>
Entity2	<p>La segunda entidad de un par de entidades extraídas a partir del texto de entrada.</p>
Entity2 Type	<p>El tipo de entidad de la segunda entidad del par de entidades extraídas a partir del texto de entrada.</p> <p>El tipo de entidad es uno de los siguientes:</p> <ul style="list-style-type: none"><li>• <i>Person</i></li><li>• <i>Organization</i></li><li>• <i>Location</i></li></ul>

# Notices

© 2017 Pitney Bowes Software Inc. Todos los derechos reservados. MapInfo y Group 1 Software son marcas comerciales de Pitney Bowes Software Inc. El resto de marcas comerciales son propiedad de sus respectivos propietarios.

### *Avisos de USPS®*

Pitney Bowes Inc. posee una licencia no exclusiva para publicar y vender bases de datos ZIP + 4® en medios magnéticos y ópticos. Las siguientes marcas comerciales son propiedad del Servicio Postal de los Estados Unidos: CASS, CASS Certified, DPV, eLOT, FASTforward, First-Class Mail, Intelligent Mail, LACS<sup>Link</sup>, NCOA<sup>Link</sup>, PAVE, PLANET Code, Postal Service, POSTNET, Post Office, RDI, Suite<sup>Link</sup>, United States Postal Service, Standard Mail, United States Post Office, USPS, ZIP Code, y ZIP + 4. Esta lista no es exhaustiva de todas las marcas comerciales que pertenecen al servicio postal.

Pitney Bowes Inc. es titular de una licencia no exclusiva de USPS® para el procesamiento NCOA<sup>Link</sup>®.

Los precios de los productos, las opciones y los servicios del software de Pitney Bowes no los establece, controla ni aprueba USPS® o el gobierno de Estados Unidos. Al utilizar los datos RDI™ para determinar los costos del envío de paquetes, la decisión comercial sobre qué empresa de entrega de paquetes se va a usar, no la toma USPS® ni el gobierno de Estados Unidos.

### *Proveedor de datos y avisos relacionados*

Los productos de datos que se incluyen en este medio y que se usan en las aplicaciones del software de Pitney Bowes Software, están protegidas mediante distintas marcas comerciales, además de un o más de los siguientes derechos de autor:

© Derechos de autor, Servicio Postal de los Estados Unidos. Todos los derechos reservados.

© 2014 TomTom. Todos los derechos reservados. TomTom y el logotipo de TomTom son marcas comerciales registradas de TomTom N.V.

© 2016 HERE

Fuente: INEGI (Instituto Nacional de Estadística y Geografía)

Basado en los datos electrónicos de © National Land Survey Sweden.

© Derechos de autor Oficina del Censo de los Estados Unidos

© Derechos de autor Nova Marketing Group, Inc.

Algunas partes de este programa tienen © Derechos de autor 1993-2007 de Nova Marketing Group Inc. Todos los derechos reservados

© Copyright Second Decimal, LLC

© Derechos de autor Servicio de correo de Canadá

Este CD-ROM contiene datos de una compilación cuyos derechos de autor son propiedad del servicio de correo de Canadá.

© 2007 Claritas, Inc.

El conjunto de datos Geocode Address World contiene datos con licencia de GeoNames Project ([www.geonames.org](http://www.geonames.org)) suministrados en virtud de la licencia de atribución de Creative Commons (la “Licencia de atribución”) que se encuentra en <http://creativecommons.org/licenses/by/3.0/legalcode>. El uso de los datos de GeoNames (según se describe en el manual de usuario de Spectrum™ Technology Platform) se rige por los términos de la Licencia de atribución. Todo conflicto entre el acuerdo establecido con Pitney Bowes Software, Inc. y la Licencia de atribución se resolverá a favor de la Licencia de atribución exclusivamente en cuanto a lo relacionado con el uso de los datos de GeoNames.



3001 Summer Street  
Stamford CT 06926-0700  
USA

[www.pitneybowes.com](http://www.pitneybowes.com)