

# Analyse de textes

## Comment trouver des informations sur un événement dans un texte du Web ?

Prendre un texte sur le Web



Remplir automatiquement un formulaire d'information sur un événement donné

Événement:	<b>séisme</b>
Lieu:	<b>Oran</b>
Date:	<b>vendredi soir</b>
Magnitude:	<b>5,5</b>
Heure:	
Dégâts:	



Où trouver le texte contenant l'information dans la page ?

Utilisation de la structure de la page et repérage des zones avec beaucoup de texte

Un séisme d'une magnitude de 5,5 degrés sur l'échelle ouverte de Richter a été enregistré **vendredi soir** dans la région d'Oran (430 km à l'ouest d'Alger), a annoncé la radio publique.

Oran, la métropole de l'ouest algérien, avait déjà été secouée en **janvier** par un **tremblement de terre** d'une magnitude de 5,3 sur l'échelle de Richter. Cette secousse n'avait fait ni victime, ni dégâts.

[...]  
L'Algérie, dont le nord est situé dans une zone sismique, est régulièrement affectée par des **tremblements de terre**.

Alger et sa région avaient été frappées, le **21 mai 2003**, par un violent **séisme** qui avait fait **2.300 morts et plus de 10.000 blessés**.

Comment séparer les différentes informations présentes ?

Utilisation des mots et expressions liées au temps (dates, temps des verbes) pour découper le texte

Un **séisme** d'une magnitude de **5,5** degrés sur l'échelle ouverte de Richter a été enregistré **vendredi soir** dans la région d'Oran (430 km à l'ouest d'Alger), a annoncé la radio publique.

**Oran**, la métropole de l'ouest algérien, avait déjà été secouée en **janvier** par un **tremblement de terre** d'une magnitude de **5,3** sur l'échelle de Richter. Cette secousse n'avait fait **ni victime, ni dégâts**.

[...]  
L'**Algérie**, dont le nord est situé dans une zone sismique, est régulièrement affectée par des **tremblements de terre**.

**Alger** et sa région avaient été frappées, le **21 mai 2003**, par un violent **séisme** qui avait fait **2.300 morts et plus de 10.000 blessés**.

Événement  
Lieu  
Date

Heure  
Magnitude  
Dégâts

Comment trouver les éléments importants dans le texte ?

Analyse grammaticale du texte et utilisation de dictionnaires spécialisés

Un **séisme** d'une magnitude de **5,5** degrés sur l'échelle ouverte de Richter a été enregistré **vendredi soir** dans la région d'Oran (430 km à l'ouest d'Alger), a annoncé la radio publique.

**Oran**, la métropole de l'ouest algérien, avait déjà été secouée en **janvier** par un **tremblement de terre** d'une magnitude de **5,3** sur l'échelle de Richter. Cette secousse n'avait fait **ni victime, ni dégâts**.

[...]  
L'**Algérie**, dont le nord est situé dans une zone sismique, est régulièrement affectée par des **tremblements de terre**.

**Alger** et sa région avaient été frappées, le **21 mai 2003**, par un violent **séisme** qui avait fait **2.300 morts et plus de 10.000 blessés**.

Événement principal  
Événement secondaire  
Contexte

Remplir automatiquement un tableau récapitulatif avec les informations extraites de toutes les pages Web récupérées

	Date Article	Titre	DATE	TIME	LOCATION	MAGNITUDE	DAMAGES
2	2008-06-07	Algérie: séisme de 5,5 degrés à Oran. L...	2008-06-06	21:02:00 local	Alger	5,5	11 blessés
1	2008-06-08	Chine: forte réplique dans la zone du L...	2008-06-08	10:51:00 UTC	Pékin	5	glissements de terrain
1	2008-06-08	Grèce: un mort dans le violent séisme d...	2008-06-08	12:30:00 UTC	Athènes	6,5	un mort
2	2008-06-08	Séisme de magnitude 5,1 en Chine	2008-06-09		Chine	5,1	glissement de terrain
2	2008-06-12	Une secousse de magnitude de 4 ressent...	2008-06-12		Liban	4	pas de dégâts
2	2008-06-14	Violent séisme dans le nord du Japon. 3...	2008-06-14	23:43:00 UTC	Japon	7	3 morts
1	2008-06-19	Séisme de 5,5 degrés de magnitude dans ...	2008-06-19	02:26:00 UTC	région pacifique du Nicaragua	5,5	