

# BBN Web Monitoring System

Distributed scalable monitoring of international Web sites



- Automatic multi-lingual data collection and mirroring of user-identified Web sites
- Automatic extraction and translation of text into English
- Search across multi-lingual sites
- Collected site archived for later use
- Translator's tool to support human translation aligned to extracted text, machine translation, and original Web page
- Publication tool for rapid assembly of products in customizable templates
- Browser-based user interface
- Designed for inter-agency data sharing

The BBN Web Monitoring System is an end-to-end capability for collecting, organizing, and translating open source content from the World Wide Web. This groundbreaking system integrates and manages the workflow of the media analysis process from beginning to end—from data collection and processing, to automated triage and retrieval, to machine-assisted translation and support for human translation, to publication and dissemination.

The system's automatic analysis of Web site content supports effective content-based retrieval and triage for human analysts who must deal with overwhelming volumes of continuously accumulating media.

**Search in English or source language!**

**Watchlist continuously monitors Web Harvesters**

Requester:	Translator:
team	languageweaver

Website	Recent Articles
Al-Hayat	681
Al-Jazeera.net English	258
allAfrica.com	80
Al-Madina	0
Al-Watan Saudi Arabia	0
Annahar	0
Aqidati	0
ArabicNews	0
Arabnews	3
Asharq Al-Awsat	68
Assafir	0
Atajdid	0

## Multi-lingual data collection

The BBN Web Monitoring System continuously captures content from user-selected Web sites into an archive that can be shared by multiple, distributed user groups. The captured site is archived and versioned for later use. Internal links are preserved in the harvested Web pages so users can navigate within the archived sites. The system also supports import of Web pages that are not automatically captured.

## Automatic translation by LANGUAGE WEAVER

Captured pages are automatically translated into English using machine translation software from Language Weaver. English speakers can use the machine translation to get the gist of an article, and then reach through the network for human translation support when a high quality translation is desired.

The BBN Web Monitoring System can support any language currently available from Language Weaver, including Arabic, Farsi/Persian, Mandarin Chinese, Russian, Korean, Somali, Hindi, and many others.

## Synchrony across all forms of data

Users can display four views of a captured Web page:

- Original Web page (including graphics, pictures, reader comments and blogs, and browseable intra-site links)
- Extracted text in the original language
- Machine translation
- Human translation, if one has been produced for the page

All forms of the Web page are aligned across passages—clicking on a passage of text in any version aligns the other versions to the same content on the screen.



Equivalent passages are highlighted

Machine translation

Extracted text

Site harvested from source

Browseable links in mirrored site

## Translingual search and triage

Users can search the system's archives in English and get results from sites in any language being harvested by the system. Personalized watchlists continuously monitor the archives for targeted content, alerting users when the system captures a relevant page. The system displays a list of pages that contain the search terms, and users can jump directly to those pages for a more thorough review.

## Specialized tools to support collaboration

The BBN Web Monitoring System includes specialized tools for triage, translation, and publication. Communication among users is extremely simple:

- From the Triage tool, analysts request human translations from a linguist with a click of a button, with the system handling all workflow between the users.
- From Translator's Aide, linguists produce a translation that is closely synchronized with the original source Web page, and return their work to the analyst when it is ready for review.
- Teams of analysts can include both Web source material and their own original content in publications, selecting relevant passages for inclusion without having to deal with formatting or editing operations.
- The Publication tool allows editors to quickly produce dissemination-ready products by collecting analyst contributions and rendering them with a fully customizable template.

BBN WMS Web Monitoring System

Triage Translate Import Publish System

team ?

Submit Cancel MT

**Taliban claims responsibility for bombings Kandahar**

The Taliban forces letter considers blasts NATO (AFP-archive)

Afghan Taliban Movement hurried to adopt a series of explosions that rocked Kandahar, south of the country yesterday, killing dozens, and considered it a letter addressed to the forces of the North Atlantic Treaty Organization ( NATO), which plans to launch an attack there.

**Taliban claims responsibility for bombings Kandahar**

The Taliban forces letter considers blasts NATO (AFP-archive)

Afghan Taliban Movement hurried to adopt a series of explosions that rocked Kandahar, south of the country yesterday, killing dozens, and considered it a letter addressed to the forces of the North Atlantic Treaty Organization ( NATO), which plans to launch an attack there.

She said the Taliban in a statement on its website that it carried out the bombings in the city, "as a message" to the leaders of the North Atlantic Treaty Organization who are planning a major operation there, adding that it has inflicted heavy losses "the enemies of the mujahedeen."

At least 30 people were killed and 52 wounded in the explosions that Afghan authorities said that "suicide bombers" from the Taliban launched in parts of the city of Kandahar, amid expectations that the number of victims.

The interior ministry spokesman said **zmaaa [زمراي]** **Bashan** that a series of explosions took place in central Kandahar Saturday night, and that one of them targeted the central prison, followed by firing missiles on the city.

In the local police officer said that five explosions took place in different parts of Kandahar by bicycles **wanyah [وناريه]**.

The targeting of prison

Officials said that the biggest attack that targeted Kandahar prison, which is located on the outskirts of

**طالبان تتبنى تفجيرات قندهار**

طالبان تعتبر التفجيرات رسالة لقوات الناتو (الفرنسية-أرشيف)

سارعت حركة طالبان الأفغانية إلى تبني سلسلة التفجيرات التي هزت قندهار جنوب البلاد أمس وأودت بحياة العشرات، واعتبرتها رسالة موجهة إلى قوات حلف شمال الأطلسي ( ناتو ) التي تخطط لشن هجوم هناك.

وقالت طالبان في بيان على موقعها على الإنترنت إنها نفذت التفجيرات في المدينة "كرسالة" إلى قادة حلف شمال الأطلسي الذين يخططون لعملية كبيرة هناك، مضيفة أنها ألحقت خسائر فادحة "الأعداء المجاهدين".

وقتل 30 شخصا على الأقل وأصيب 52 آخرون في التفجيرات التي قالت السلطات الأفغانية إن "انتحاريين" من طالبان شنوها في أنحاء مدينة قندهار، وسط توقعات بارتفاع عدد الضحايا.

وقال المتحدث باسم وزارة الداخلية **زمراي بشاري** إن سلسلة تفجيرات وقعت بوسط قندهار مساء السبت، وإن أحدها استهدف السجن المركزي وتلاه إطلاق صواريخ على المدينة.

وقال ضابط في الشرطة المحلية إن خمسة انفجارات وقعت في أنحاء مختلفة من قندهار بواسطة دراجات هوائية ونارية.

استهداف السجن

وقال مسؤولون إن أكبر هجوم استهدف سجن قندهار الذي يقع على مشارف المدينة، في محاولة -على ما يبدو- لتكرار عملية

Human translation created here

Copy and edit or enter original translation

Passages aligned across views

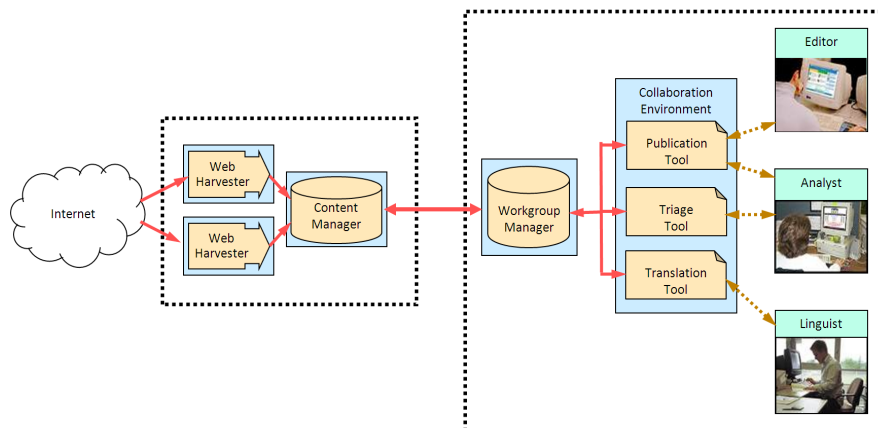
People, places, and organizations are highlighted and linked to the transcript.

## Inter-Agency Data Sharing

The BBN Web Monitoring System is designed as a distributed Internet system that allows its components and services to be optimally located near its media sources and human users. The distributed design supports easy expansion, robust failure recovery, and inter-agency data sharing. The system consists of the following components:

- One or more Web Harvesters plus a Content Manager, which can be located at any facility with sufficient infrastructure to support them. Each Web Harvester supports a single language. These components are intended to be maintained by ISP professionals and may be shared by multiple, distributed user groups.
- A Workgroup Manager, which is located in each end user's environment to support a local, private, searchable archive of products. The Workgroup Manager includes the local database and the Collaboration Environment, which supports the user tools for triage, translation, and publication.

The Collaboration Environment tools are all browser-based and designed to work with Microsoft Windows (XP and Internet Explorer) and require no additional custom software to be loaded on the client. In addition, the system's single browser-based interface eliminates the need for users to learn and maintain a variety of tools.



## Federated Search

Multiple Web Harvesters can be deployed to support different languages and target Web sites and are connected to one or more Content Managers. Each Workgroup has its own private database where products are stored, but a Workgroup can access the shared databases of any number of Content Managers via the system's federated search capability.

For more information about the BBN Web Monitoring System and other advanced media exploitation technologies, please call or write:

Raytheon BBN Technologies  
10 Moulton Street  
Cambridge, MA 02138  
617.873.2522  
web-monitoring@bbn.com  
www.bbn.com

**Raytheon**  
**BBN Technologies**