

Past TIPSTER Research at NMSU

Architecture Committee

Document Manager Implementation

Document Manager Validation Suite

Multi-Lingual Text Retrieval

Graphical User Interface Development

Multilingual Named Entity Task

Oleada Demonstration Project

Corelli Demonstration Project

TIPSTER Software Available from NSMU

TDM - TIPSTER document manager

Implements all "document manager" functions

TIPSTER Validation suite

An architecture compliance checker

TUIT - TIPSTER User Interface Toolkit

Collection of GUI modules for viewing and managing TIPSTER objects (e.g. Documents and Annotations)

TR - Tabula Rasa

Now TIPSTER capable MUC template tool constructor

Annotators

Conversion of MUC named entity and Juman and Chinese segmentation to annotations.

Available on WWW <http://crl.nmsu.edu/Tools/Software>

TIPSTER Phase III Research at NMSU

Unicode and Multilingual TIPSTER Detection

Visualizing Multilingual Collections

Analyze TIPSTER Technology User Needs

Analyze TIPSTER Technology User Needs

Identify and observe government TIPSTER
technology users

Design scenarios based on user observations

Identify or design experimental prototype GUIs

Conduct research and evaluation of iterative prototype
refinements

Report and apply lessons learned to our development of
multilingual collection browsing and query development
tools

Unicode and Multilingual TIPSTER Detection

Can Unicode be used in a TIPSTER-compliant detection system?

How can Unicode codeset conversion, segmentation and morphology modules be folded into the TIPSTER architecture?

Can existing detection systems be expanded to use Unicode?

Can a modular query-translation system be implemented in the TIPSTER framework?

Can Unicode be used in a TIPSTER-compliant detection system?

Yes.

Unicode has many advantages for detection tasks:

Systematic handling of most languages in 46,000 defined codepoints.

Rules for compositional and non-compositional characters and their normalization.

Always logical order (typing order) in the byte stream.

Display technology is available and rapidly improving at CRL.

How can Unicode codeset conversion, segmentation and morphological analysis modules be folded into the TIPSTER architecture?

Codeset conversion modules can be applied to imported TIPSTER documents and collections.

Segmentation and morphological analysis are best handled by Unicode-savvy TIPSTER annotators.

We will develop Unicode annotators for Chinese and Spanish segmentation and morphological analysis.

Can existing detection systems be expanded to use Unicode?

Probably.

We will evaluate several detection systems to determine how to incorporate them into a TIPSTER Unicode framework.

Keeping segmentation and morphology as annotators should facilitate using existing engines.

Can a modular query-translation system be implemented in the TIPSTER framework?

Yes.

A query translator should operate as an annotator on a detection need, adding Unicode annotations of translation information.

Multiple languages can be handled in a language-independent manner with Unicode annotations.

Transforming a DetectionNeed into a DetectionQuery requires accessing the correct annotation set.

Visualizing Multilingual Collections

How can large multilingual document collections be usefully visualized?

How can multilingual queries be visualized in relationship to collections?

How can visualization tools be integrated into a TIPSTER architecture?

How can large multilingual document collections be usefully visualized?

Using Pathfinder networks that show the “essential” relationships between collection terms.

By operating on Unicode TIPSTER document collections that have segmentation and morphology markup.

Designed to meet the needs of the visualization system consumer.

How can multilingual queries be visualized in relationship to collections?

By highlighting and emphasizing terms drawn from a Unicode DetectionNeed in a collection Pathfinder network.

By identifying the most statistically significant terms that cooccur with query terms in a document collection and drawing Pathfinder networks for these terms.

By using DetectionNeed Unicode translation annotations to identify and emphasize terms in the collection network.

How can visualization tools be integrated into a TIPSTER architecture?

By extracting cooccurrence statistics for Pathfinder networks from a Unicode DocumentCollectionIndex.

By using the DetectionNeed and Unicode annotations to extract query information.