

European digital content for the global networks

EDC-22071 KOM2002/27938

Groupware for Distributed Content Production

KOM2002

Access to the online translation engine

Security (distribution level)	Public
Contractual date of delivery	31.12.2002
Actual date of delivery	30/9/2003
Deliverable number	D 8.1
Deliverable name	Access to the online translation engine
Туре	PP
Status & version	Final report
Number of pages	
WP contributing to the	WP8
deliverable	
WP / Task responsible	SYSTRAN
Other contributors	
Author(s)	Elsa Sklavounou
EC Project Officer	Kimmo Rossi
Keywords	Advanced Protocole Interface, Translation Engines
Abstract (for dissemination)	This report aims to present the access given to the Web4Haelth interface
	for testing, evaluation and integration of the MT component

TABLE OF CONTENTS

1. SUMMARY	2
Normalisation tool	
Extraction tool	
Encoding tool	
2. PRESENTATION	3

1. Summary

Normalisation tool

Pre-filters have been constructed according to the authentic feedback given by NetDoctor.

SYSTRAN believes that the FAQs produced by medical partners is gold and do not show the real status of questions that patients will pause in an online chat and forum session. The normalisation tool is ready to process any agrammatical input to render it translatable. Once set and implemented, the normalisation tool will be maintained and process in that way the most frequent and systematic input agrammaticalities.

Extraction tool

The developed extraction tools allow us to extract accurate terminology from the gold parallel corpus that medical partners have produced in all the language pairs of the project. The entries will be formatted in a form of bilingual dictionary that will be automatically processed by the encoding tool in order to be taken into account during the translation process.

Encoding tool

The encoding tool has been revised for the specific needs of KOM2002 in the language pairs of the project. The experts are able to introduce bilingual entries in EN, FR, EL, SV, DE to improve the translation results via a terminological workbench. The encoding tool serves to give a minimal morphosyntactic encoding rendering grammatically possible the application of those entries in medical contexts.

English to Swedish prototype

The developed prototype English to Swedish renders possible the introduction of English to Swedish bilingual dictionaries. At the present stage DNTs (Do Not Translate words) in English can be taken into account during the translation process from English to Swedish.

Evaluation tool

A specifically defined version of SYSTRAN Review Manager (SRM), which is a productive tool, used for reviewing, quality assessment and maintenance of linguistic resources is being under elaboration. Medical experts will be able to review the translation results and improve the translation quality during evaluation by the introduction of more accurate terminology.

API

The Advanced Protocol Interface is a turnkey website translation solution. It transforms standard websites and content applications into interactive multilingual hubs automatically.

2. Presentation

The Advanced Protocol Interface is the more convenient way to give access to the online translation engines in a Web-based groupware for distributed content production like Web4Health.

An API account has been created for KTH, which will host the Web4Health platform. SYSTRAN gave to KTH the means to configure and integrate the translation engines according to the ergonomy of the functional interface. Access to the translation engines in all language pairs has been provided to KTH in order to test, evaluate and implement the language pair translation engines according to the needs and the terms of the KOM2002 project.

The API is directly linked to the terminological workbench. The users/content providers can customize the machine translation output integrating online the KOM2002 specialised terminology. The KOM2002 external dictionary can be updated, modified and maintained by users/content providers under the control of the website administrator. The updated version of the external dictionary is automatically taken into account in every translation process.

The identification of SYSTRANLinks API kom2002-api is now associated to the following user account

```
email | passwd
-----
jpalme@dsv.su.se | orion1
```

Login on http://www.systranlinks.com/

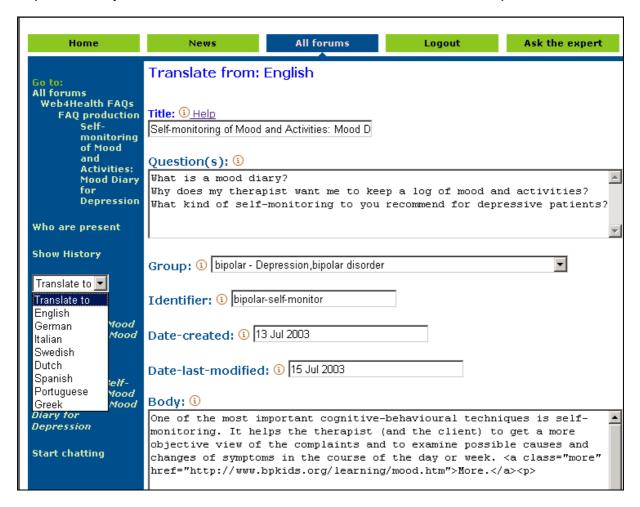
SYSTRAN provided technical support to implementation issues

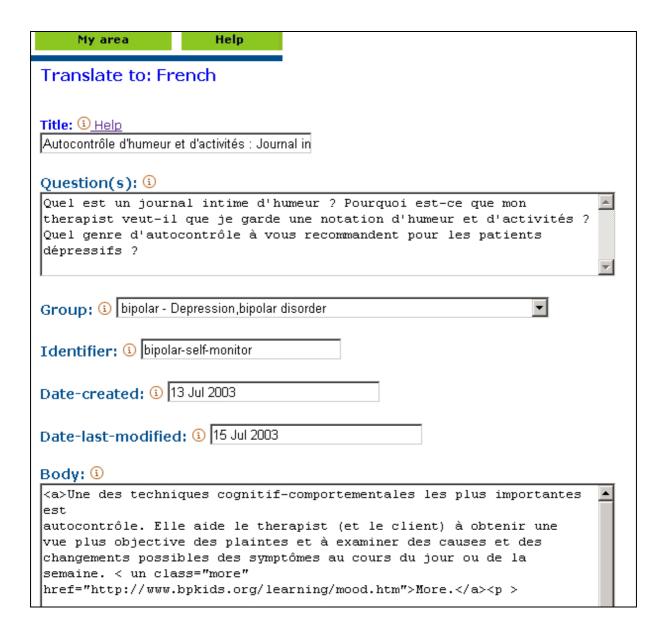
```
> POST /trans HTTP/1.0
> Connection: close
> Content-type: multipart/form-data;
> boundary=-----69819914916652
> Content-Length: 635
> -----69819914916652
> Content-Disposition: form-data; name="noxml"
> 1
> -----69819914916652
> Content-Disposition: form-data; name="lp"
> en de
> -----69819914916652
> Content-Disposition: form-data; name="api"
> 1
> -----69819914916652
> Content-Disposition: form-data; name="text"
> This message is written in English.
> Will it be translated to German?
> I hope it will succeed.
> -----69819914916652
> Content-Disposition: form-data; name="id"
> kom2002-api
> -----69819914916652--
>
>>
> HTTP/1.1 302 Found
> Date: Fri, 14 Feb 2003 19:00:20 GMT
> Server: Apache/1.3.23 (Unix) mod_perl/1.26 mod_ssl/2.8.6 OpenSSL/0.9.6
> Location: http://www.systranet.com/
> Connection: close
> Content-Type: text/html; charset=iso-8859-1
```

Due to HTTP protocol description directly written to the socket of the server. To make sure that KTH request complies with the proper RFC, SYSTRAN highly recommended that KTH uses some standard client library such as http://jakarta.apache.org/commons/httpclient/api/index.html

Most issues like the ones described come from improper HTTP protocol, and since this is a pretty complex matter, it is simpler to use a well-known library which encapsulates all that.

SYSTRAN Machine Translation Services are implemented and being used experimentally for the translation of chats, discussion forums, FAQs production:

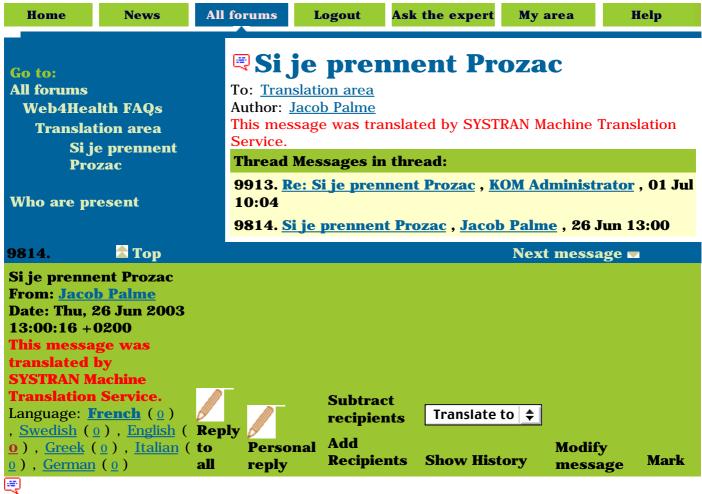




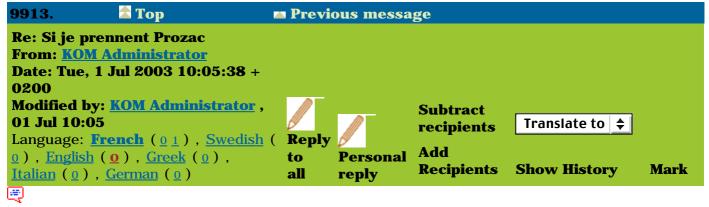
My area	Help
Translate to: Sw	edish
Title: 1 Help	
Self-monitoring av moode	en och aktiviteter:
Question(s): 1	
Är vad en mooddiar hålla en journal s av self-monitoring	v mooden och
Group: (1) bipolar - De	pression,bipolar
Identifier: (1) bipolar	self-monitor
Date-created: (1) 13	Jul 2003
Date-last-modified	15 Jul 2003
Body: 🗓	
<a>En av de viktig self-monitoring. I mer mål sikt av kl ändringar av tecke href="http://www.k	en hjälper t agomålen och n i kursen a



Si je prennent Prozac 03-08-12 12:36



Je suis malheureux en hiver. Si je prennent Prozac.



Les traductions seront-elles meilleures avec le point d'interrogation ? Je suis malheureux en hiver. Au cas où je prendre Prozac ?

Home News All forums Logout Ask the expert My area Help
You are logged in as Jacob Palme
Today's date: Tue, 12 Aug 2003 12:35:41 +0200

Ich wenn, nehmen Prozac 03-08-12 12:36



Ich bin im Winter unglücklich. Ich wenn, nehmen Prozac.



Sind Übersetzungen mit dem Fragezeichen besser? Ich bin im Winter unglücklich. Sollte ich Prozac nehmen?

Home News All forums Logout Ask the expert My area Help
You are logged in as Jacob Palme
Today's date: Tue, 12 Aug 2003 12:36:02 +0200

Bör jag, tar Prozac 03-08-12 12:36



Jag är unhappy i winteret. Bör jag, tar Prozac.



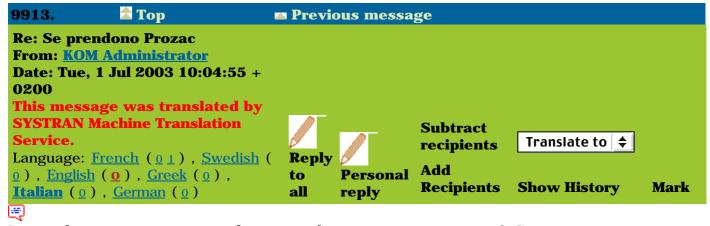
Är översättningar bättre med frågefläcken? Jag är unhappy i winteren. Bör jag ta Prozac?

Home News All forums Logout Ask the expert My area Help
You are logged in as Jacob Palme
Today's date: Tue, 12 Aug 2003 12:36:30 +0200

Se prendono Prozac 03-08-12 12:37



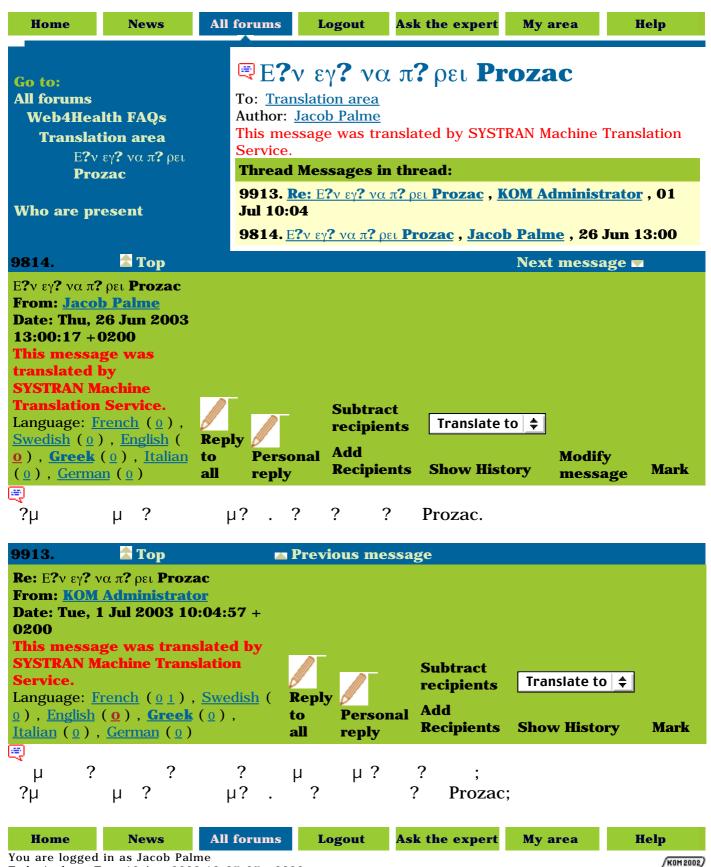
Sono infelice nell'inverno. Se prendono Prozac.



Le traduzioni saranno migliori con il punto interrogativo? Sono infelice nell'inverno. Dovrei prendere Prozac?

Home News All forums Logout Ask the expert My area Help
You are logged in as Jacob Palme
Today's date: Tue, 12 Aug 2003 12:36:48 +0200

??? ??? ?? Prozac 03-08-12 12:37



Today's date: Tue, 12 Aug 2003 12:37:07 +0200