

STIG, System for Translation Information in General: User manual and documentation

Diana Santos

2nd version, 30 October 2017 (1st version 4 June 2017)

This is on purpose an attempt to write for people who know very little computer science, and the reason is that it is to be (hopefully) used – and or improved – by future research assistants, or other colleagues who have not been involved in the design process. It will also serve as a documentation on how the instantiation of STIG for PANTERA was created.

Basics about the technology

The idea is describe a “universe” where there are several kinds of entities, each with different properties which are relevant, and not a main one (as it is, simplifying somehow, the case of traditional databases, where you have a single entity, say books, or customers). In addition there may be several relationships between entity types, both of the same kind (like a book may be inspired by another book, or a person may be married to another person, and of different kinds (a book is written by a person). And the same two specific entities (tokens) can have several different relationships between them (like book1 was written by person1, and book1 was edited by person1).

One of way of describing this kind of “information universes” is to use topic maps. It is not the implementation that concerns us here (topic maps are usually implemented on top of ... databases!), but the expression capabilities that the technology provides. You can think of this as several levels of analysis of the same book: physically, it is just letters on paper, but you can provide several different networks and entities and objects to describe it (and in principle there is an infinite number of possible analyses).

Let us use the concrete example of PANTERA-related information, to describe translation between Portuguese and Norwegian, to see how one would go about developing such topic-map system. The URL is: <https://stig.hf.uio.no/pantera/> where anyone can search the information, but note that one has to have administration (“admin”) privileges to design the contents and fill in the data.

The first thing one needs to decide is which entity types is one going to care about.

Well, in fact, PANTERA is bilingual, so there are two languages (specified by language codes) one needs to consider, namely Portuguese and Norwegian. Working languages is what you have to define first, and they will be the languages of the interface(s). The administration language is English, as you can see in the header and in all the subsequent figures. (One can, in fact, change that as well by using the **Sysvars** menu in the administration mode, but we will leave this for later, because it helps our teaching purposes here.) In Figure 1, we define our working languages.

Language	Default admin	Default public	Delete
Norwegian	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Portuguese	<input type="radio"/>	<input checked="" type="radio"/>	<input type="checkbox"/>

Buttons: Add language, Save changes

Figure 1 Defining "languages" in STIG

Defining entity types in STIG

In the practical context of the STIG system (administrative part), you select **Entity Types**, and define them by giving them names (in several languages, if you are a bi- or multilingual developer) and describe their relevant properties.

As you see in Figure 2, you start by providing the following information: the name in the working languages (Name (por) and Name (nob)) and its (optional) description, also in both languages (Description (por) and Description (nob)). The order has to do with the presentation of the different entity types in other contexts, and it is chosen by the user.

And for each entity type, one can edit or delete it entirely. (Or create a new one, of course.)

Edit Entity (ID 1)

Name (por):

Name (nob):

Description (por):

Description (nob):

Order:

Has static content that can be displayed on splash screen

Buttons: Oppdater, Delete

Registered properties: Nome/Navn, Nasceu/Født, Morreu/Ødt, País de origem/Opprinnelsesland [add, edit, remove](#)

Edit Entity (ID 52)

Name (por):

Name (nob):

Description (por):

Description (nob):

Order:

Has static content that can be displayed on splash screen

Buttons: Oppdater, Delete

Registered properties: identificador/id, páginas no original/sidene i kilde, páginas da tradução/sidene i oversettelse [add, edit, remove](#)

Figure 2 Defining entity types in STIG

Take, for example, the action of Edit Entity (ID 51) (editing the entity type Antologia/Antologi). The properties that this particular entity type, representing an anthology, has, are displayed in the bottom line in a smaller font, together with link to add, edit or remove them:

Registered properties: Título/Tittel, Responsável/Ansvarlig, Língua/Språk [add, edit, remove](#)

In this particular case, and translating the property names into English, we can add, edit or remove the title of the anthology, language of publication, responsible for the choice(s). Or, obviously, create new ones.

An overview of all the entity types described so far for PANTERA is (names in Portuguese/Norwegian, followed by an English equivalent. The English is just for this report). The ID numbers are given by the system.

- 1 Autor/Forfatter (author)
- 45 Tradutor/Oversetter (translator)
- 23 Obra/Verk (work)
- 46 Tradução/Oversettelse (translation)
- 24 Publicação/Utgivelse (publication)
- 43 Editora/Forlag (publisher)
- 44 Coleção/Serie (collection)
- 51 Antologia/Antologi (anthology)
- 47 Prémio/Pris (prize)
- 48 Atribuição de prémio/Tildelte priser (prize attribution)
- 49 Corpo/Korpus (corpus)
- 52 Excerto num corpo/korpusekstrakt (corpus excerpt)
- 50 Língua/Språk (language)
- 30 Pessoa/Person (person)
- 495 Metadados/Metainformasjon(test) (metadata)

For each of these entities, we define at the same time (or later) the set of properties or attributes that they are supposed to have.

For example, for Editora/Forlag, it shows

Registered properties: Fundada/Grunnlagt, Sede/Hvor, Nome/Navn

Properties for entity "Publicação/Utgivelse"

Figure 3 Using the "Properties" menu

Back to Publicação/Utgivelse and editing it, we can see in Figure 3 that also each property has its own id, and that there are a lot of things one can specify about each property, see screenshot 1, that shows properties data/dato (date) and Local de publicação/Publiseringssted (publication place), with respectively identifiers id37 and id38.

Not everything is equally relevant at first sight, but it is important to note that one can go there and change and improve the properties; it is not required to do it all at once.

Some of this information concerns its semantics, such as the name and the type of values it expects. If numbers, what kind of numbers; if dates, which granularity; if names, do they belong to a closed set, or can be any string. Other information is about how to interpret what is put there: are the strings in HTML, are the numbers geographical coordinates, are they supposed to be presented grouped in intervals or directly searchable, etc.

For some kinds of values, it is possible to automatically validate them, specifying a validator and even specific error messages so that a user can understand what went wrong.

Only after the entity types have been characterized by properties can we allow entities of that type to populate the system. In the next table we present the properties that have been defined for PANTERA so far:

Entity type	Properties
-------------	------------

Autor / Forfatter	Nome/Navn (id1), País de origem/Opprinnelsesland (id28), Nasceu/Fødselsdato (id26), Morreu/ Dødsdato (id27)
Tradutor/ Oversetter	Nome/Navn (id40), Data de nascimento/ Fødselsdato (id41), Data de falecimento / Dødsdato (id42), Nacionalidade / Nasjonalitet (id43), Língua materna / Morsmål (id44)
Obra/Verk	Título/Tittel (id29), Local de escrita / Sted der verket ble skrevet (id30), Intervalo de escrita / Periode da verket ble skrevet (id39), Género/ Sjanger (id65), Autor/Forfatter (id81)
Tradução/ Oversettelse	Língua/Språk (id45), Título/Tittel (id46), Período da tradução/ Oversettelsesperiode (id47), Língua imediata/ Siste mellomspråk (id67), Tradutor/Oversetter (id75), Género/ Sjanger (id84)
Publicação/ Utgivelse	Data/Dato (id37), Local de publicação / Publiseringssted (id38), Original ou Tradução / Kilde eller oversettelse (id48), Tiragem / Antall eksemplarer (id49), Número da edição / Utgave nummer (id50), ISBN/ISBN (id56), Publicação digital / Digital publisering (id57), Número de páginas / Sidetall (id66), Capa/Forside (id77), Título/Tittel (id78), Autor/Forfatter (id79), Editora/Forlag (id80)
Editora/ Forlag	Nome/Navn (id51), Fundada/Grunnlagt (id32), Sede/Hvor (id33)
Coleção/ Serie	Nome/Navn (id34), Responsável/Redaktør (id35), Tamanho/Antall (id36), Língua/Språk (id82), Data do primeiro volume/ Første bind (id83)
Antologia/ antologi	Título/Tittel (id62), Responsável/Ansvarlig (id63), Língua/Språk (id64)
Prémio/Pris	Nome/Navn (id51)
Atribuição de prémio / Tildelte priser	Data/Dato (id52), Que prémio/Hvilket pris (id74)
Corpo/ Korpus	Nome/Navn (id54), URL/URL (id55)
Excerto num corpo /Korpusekstrakt	Identificador/Id (id68), Páginas no original/Sidene i kilde (id69), Páginas da tradução/Sidene i oversettelse (id70)
Língua/ Språk	Nome/Navn (id73)
Pessoa/ Person	Nome/Navn (id59), Data de nascimento/ Fødselsdato (id60), Data de falecimento / Dødsdato (id61)
Metadados /Metainformasjon(test)	Conteúdo/Innhold (id72)

Note that different entity types can have similar properties, and also that properties names can also be named just like entity types (see the Língua/Språk). This may confuse us, but not the system.

Another extremely relevant information that we want to be able to specify is relationships between these kinds of entities, and it is here that the full power of topic maps shines, because one can name (and record) several different relationships (and their inverses) among the same (kinds of) entities. These are called Associations and have been deemed so important that are supposed to be defined even before the actual properties of the entities, as you can see in STIG's header in administration mode.

As should by now be expected, also associations have a different id, and are described in both languages of the interface. See Figure 2 for a screenshot defining a relationship between an author and a work (the description is merely for documentation purposes).

STIG Languages > Entity types > **Associations** > Properties > Field Formats > Import > Values > Layout element >

dssantos log out

Subject type

Autor/Forfatter

Association Name (por)

escreveu

Association Name (nob)

skrev

Association Name (por)

cujo autor é

Association Name (nob)

hvis forfatter er

Object type

Obra/Verk

Description (por)

relação entre o autor e a obra

Description (nob)

mellom forfatter og verk

Save
Delete

Edit Association (id 29)

Subject type

Autor/Forfatter

Association Name (por)

editou na editora

Object type

Editora/Forlag

Figure 4 Using the «associations» menu

We show below all associations so far defined in PANTERA, together with their opposite associations. The system recognizes (encodes) associations by their ID, which means that we can have a different association with the same name, but between different types of entities. This is the case, for example of the (triple) association “attributed prize to”, which can relate to authors, works or translations.

Id 15	Autor/ Forfatter	escreveu/skrev	Obra/Verk	cujo autor é / hvis forfatter er
Id 29	Autor/ Forfatter	editou na editora / publiserte i forlaget	Editora/Forlag	publicou obras de / publiserte verk av
Id 39	Autor/ Forfatter	é dado como autor de / er vist som forfatter av	Tradução / Oversettelse	que tem como autor / som har som forfatter
Id 46	Autor/ Forfatter	é co-autor de / skrev sammen med	Autor/ Forfatter	é co-autor de / skrev sammen med
Id 22	Obra/Verk	foi selecionada para/ble valgt for	Excerto num corpo / Korpusekstrakt	foi retirada de / ble ekserpert fra
Id 26	Obra/Verk	trabalho foi publicado como / har blitt utgitt som	Publicação / Utgivelse	resulta de / hører til

Id 47	Obra/Verk	foi traduzida para / ble oversatt til	Tradução / Oversettelse	é tradução de / er oversettelse av
Id 20	Publicação / Utgivelse	é uma publicação de / er en utgivelse av	Obra/Verk	foi publicada como / ble utgitt som
Id 21	Publicação / Utgivelse	está diretamente relacionada / er relatert til	Publicação / Utgivelse	original relacionado com a tradução / kildeutgivelse relatert til oversettelse
Id 24	Publicação / Utgivelse	é uma publicação da tradução / er en utgivelse av oversettelsen	Tradução / Oversettelse	tradução foi publicada como / oversettelsen ble utgitt som
Id 38	Publicação / Utgivelse	foi publicada por / ble utgitt av	Editora/ Forlag	publicou / utga
Id 17	Coleção/ Serie	contém obras de / inkluderer verk av	Autor/ Forfatter	foi incluído em / hvis verk er inkludert i
Id 16	Tradutor/ Oversetter	traduziu o mesmo autor que / har oversatt samme forfatter som	Tradutor/ Oversetter	traduziu o mesmo autor que / har oversatt samme forfatter som
Id 18	Tradutor/ Oversetter	traduziu obras de / har oversatt verk av	Autor/ Forfatter	foi traduzido por / har blitt oversatt av
Id 19	Tradutor/ Oversetter	traduziu / har oversatt	Obra/ Verk	foi traduzida por / har blitt oversatt av
Id 37	Tradutor/ Oversetter	traduziu a mesma obra que / har oversatt samme verk som	Tradutor/ Oversetter	traduziu a mesma obra que / har oversatt samme verk som
Id 25	Tradução/ Oversettelse	é tradução de / svarer til	Obra/ Verk	foi traduzida para / ble oversatt til
Id 40	Tradução/ Oversettelse	foi publicada como / ble utgitt som	Publicação / Utgivelse	é uma publicação de / er en utgivelse av
Id 30	Antologia/ Antologi	tem obras de / omfatter verk av	Autor/ Forfatter	está incluído como autor em / er med som forfatter i
Id 32	Antologia/ Antologi	Antologia publicada por / antologi utgitt av	Editora/ Forlag	publicou a antologia / utga antologi
Id 33	Antologia/ Antologi	inclui traduções de / omfatter oversettelser av	Tradutor/ Oversetter	está incluído como tradutor em / er med som oversetter i
Id 27	Pessoa/ Person	relacionado com / er knyttet til	Editora/ Forlag	recorreu a / brukte
Id 28	Pessoa/ Person	colaborou com / samarbeidet med	Autor/ Forfatter	teve ajuda de / ble hjulpert av
Id 34	Pessoa/ Person	fez uma revisão de / anmeldte	Publicação / Utgivelse	foi recenseada por / ble anmeldt av
Id 48	Pessoa/ Person	reviu / reviderte	Publicação / Utgivelse	foi revista por / ble revidert av
Id 36	Pessoa/ Person	prefaciou / skrev forord til	Publicação / Utgivelse	foi prefaciada por / har forord av
Id 31	Pessoa/ Person	responsável por / ansvarlig for	Antologia/ Antologi	criada por / valgt ut av

Id 43	Atribuição de prémio / Prisutdeling	premiou autor / ble tildelt forfatter	Autor / Forfatter	autor premiado com / forfatter fått tildelt
Id 44	Atribuição de prémio / Tildelte priser	premiou / ble tildelt	Obra / Verk	foi premiada com / fått tildelt
Id 45	Atribuição de prémio / Tildelte priser	premiou tradução / ble tildelt oversettelse	Tradução/ Oversettelse	tradução premiada com / oversettelse fått tildelt

One can thus see that we have specified many different (possible) associations between the several entity types, and it probably gets clearer why we needed so many different entity types. Because, even though ontologically an author and a translator are also persons, by filling different roles they are different in the system, even if, say, an author is also a translator and writes a preface for another author and a review for another book, s/he will appear as a member of three different types of entities in STIG.

Note that, as far as topic maps expressivity goes, one could have defined as well the association “same person as” between translators, authors and persons, but we have not deemed it necessary in the present context.

After having defined what kinds of entities and associations among them can exist, there are two more tasks one must do: define how they are going to be presented (defining formats and layouts), and populate the system with concrete data.

Specifying formats

For each entity we can (and must) define several presentational properties: in fact, each entity type has a set of formats associated, as Figure 5 shows, and one can create other formats as well. By default, there are several formats associated to each entity type: `full display of entity`, which shows everything; `menu format`, to be used in dropdown menus; `table format`, to be used in tables, `linking`, which is the result of showing the selections in the leftmost menu, and `root`, which is used when one selects only one entity at a time.

Those formats are a combination of the values of the properties, using a simple syntax. Exemplifying with Autor/Forfatter, and knowing that an inner [] indicates the value of that property, and that an outer [] indicates if it exists, one could define all these “format elements” in the category `table format`, for example.

name => [[Navn:1]]

livstid => [[Født:26]] [- [Død:27]]

opprinnelsesland => [[Opprinnelsesland:28]]

or, in the `linking` format, when we want one line only:

Ett-felts-visning => [[Navn:1]] [[([Født:26]] [- [Død:27]]), [[Opprinnelsesland:28]]

Or, for Publicação / Utgivelse, one could define formats like this for `menu`:

Utgivelse1 => [[Tittel:78]] [- [Publiseringssted:38] ([dato:37]]

Figure 5 shows the interface to define formats:

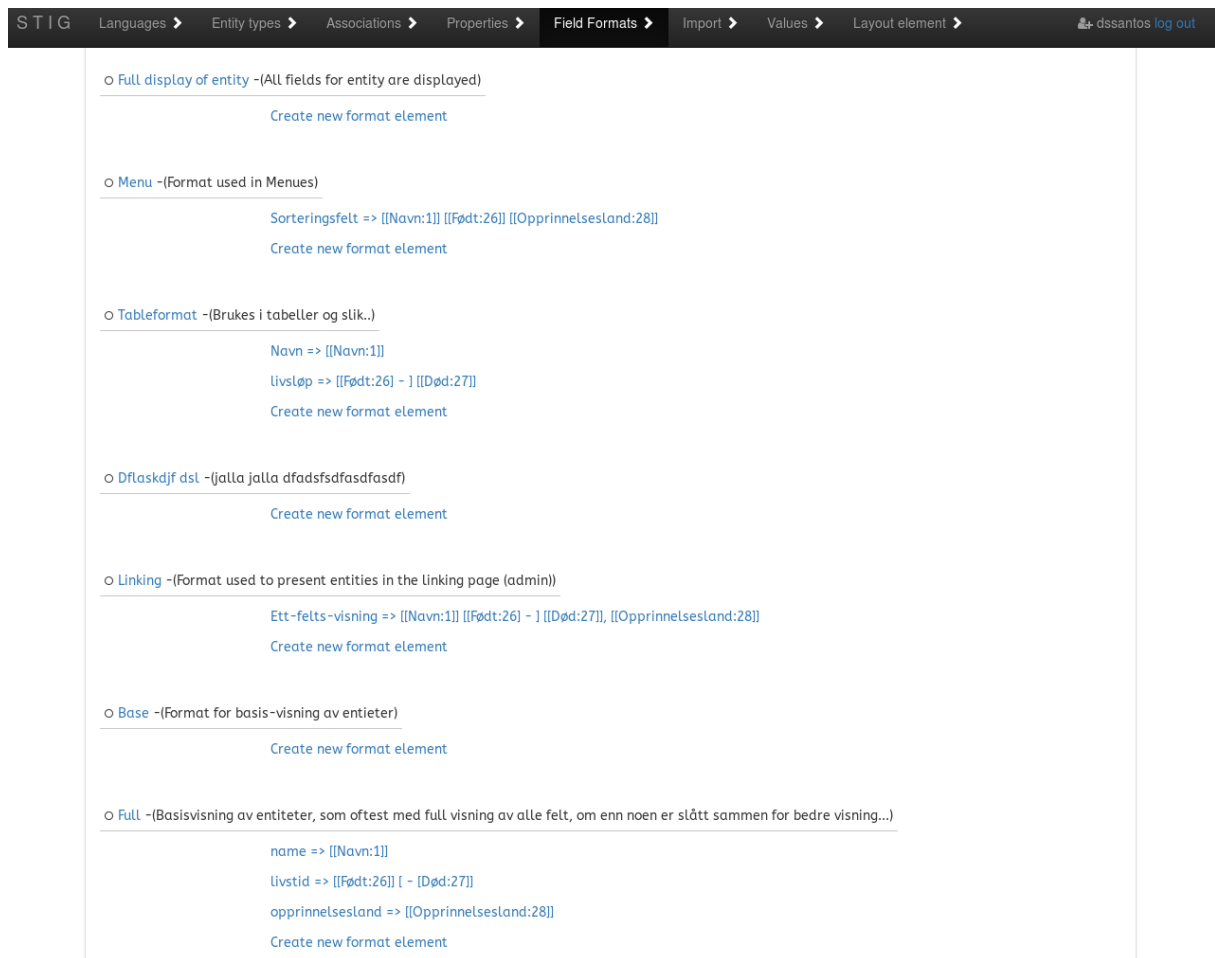


Figure 5 Defining formats in STIG

If one clicks on either the blue format elements or on the line “Create new format element”, also in blue, one is offered another interface where these format elements can be defined, and named. In fact, contrarily to all the other cases so far, STIG uses the names of the formats and of the format elements. Here there are no (at least for the user) numeric ids.

In Figure 6 you see the new interface. Alias represents the name of this format element, and you start (probably little intuitively) by giving an order property to this format element. This is in order to display the format element in the right order, vertically or horizontally, when this format is invoked. The only thing you have to care about is the relative order of the different elements. So if you have one 10 and you want this new one to become later, you choose a larger number – say 20 (or 11); if you want the new format element to come before, you write 5, or 9.

Then in Format is where you define the format element itself.

There is then an optional possibility to define some words which will not be used in sorting, like articles (this is mainly convenient for titles, but could also be used for the word “Editora”, which means publishers and is often but not consistently used in Portuguese), and also another one to truncate format output to a particular number of words.

STIG Sysvars > Languages > Entity types > Associations > Properties > Formats > Import > Values > Layouts > dssantos log out

Alias

Portuguese:

Norwegian:

Format properties

Order:

Format Element used for sorting only

Format Element invisible except for admins

Format

Portuguese:

Norwegian:

Article Handling

Portuguese:

Norwegian:

Truncation

Truncate format output

Number of words from start:

Number of words from end:

Figure 6 Defining format elements

Then there is another step in which one must define visualization properties of STIG. This is done for each entity type, for which we must define layouts. Their rationale is that we may wish to see in a different form the entities when we look at them individually, when we search for them, or when they appear as a list, or when they are shown in e.g. dropdown menus. And we may want to see them vertically or horizontally depending on different criteria... Therefore the large definition potential, which is probably the most difficult to handle by the administrator, because they are closely intertwined with formats.

Layouts thus refer to the (lower level) formats described before. In Figure 7, one can see the initial screen for an entity who has no layout yet, Corpo/Korpus.

STIG Sysvars > Languages > Entity types > Associations > Properties > Formats > Import > Values > **Layouts >** [dssantos log out](#)

Existing layouts

Base; row; Full display of entity Base Layout for Entity Use layout for Entity Splash Screen [Edit](#) [Delete](#)

[Update](#)

Create New Layout

Headline

Portuguese

Norwegian

Options

Reference Name

Display type

Format name

Class

This Layout Element creates a root menu

Limit output to related layouts (silent output)

Display Headline in layout

[Save](#)

Figure 7 Start defining layouts

But it is easier to show how to do it with an entity already with several layouts defined, namely Obra/Verk, as shown in Figure 8. One can still create more.

STIG Sysvars > Languages > Entity types > Associations > Properties > Formats > Import > Values > **Layouts >** [dssantos log out](#)

Entity Type

Existing layouts

Fullvisning; column; Full display of entity Base Layout for Entity Use layout for Entity Splash Screen [Edit](#) [Delete](#)

ListaObras; type:row; Tableformat Base Layout for Entity Use layout for Entity Splash Screen [Edit](#) [Delete](#)

Submenu; ; Menu Base Layout for Entity Use layout for Entity Splash Screen [Edit](#) [Delete](#)

Tabellvisning; row; Tableformat Base Layout for Entity Use layout for Entity Splash Screen [Edit](#) [Delete](#)

[Update](#)

Create New Layout

Headline

Figure 8 Defining layouts in STIG

When one has more than one layout, it is necessary to specify which one should be used for “base layout”. This means the layout used for inspecting one entity at the time. Also, one has to choose which of the several layouts is to be used for splash screens.

Let us edit one layout, still for Obra/Verk, in figure 9. The observant reader will note that layouts have ids again...

The screenshot shows a web interface for editing a layout. At the top, there is a navigation bar with links for 'STIG', 'Sysvars', 'Languages', 'Entity types', 'Associations', 'Properties', 'Formats', 'Import', 'Values', and 'Layouts'. The 'Layouts' link is highlighted. Below the navigation bar, there is a dropdown menu for 'Entity Type' with 'Obra/Verk' selected. The main content area is titled 'Edit Layout (id 22)'. It has two sections: 'Headline' and 'Options'. Under 'Headline', there are two input fields: 'Portuguese' with the value 'Lista de obras' and 'Norwegian' with the value 'Verklister'. Under 'Options', there are several fields: 'Reference Name' with 'ListaObras', 'Display type' with 'type:row', 'Format name' with a dropdown menu showing 'Tableformat; Brukes i tabeller og slik..', and 'Class' with 'row'. There are also three checkboxes: 'This Layout Element creates a root menu', 'Limit output to related layouts (silent output)', and 'Display Headline in Layout', all of which are unchecked. At the bottom, there is a 'Nav button ordering format' dropdown menu with the value 'Choose a format for sorting with navigational buttons (Base Layouts only)'. A blue 'Save' button is located at the bottom left of the form.

Figure 9 Editing a particular layout

Although one may not have a headline if one chooses not to (that is, if one does not tick on “Display Headline in Layout”), one must provide the text that heads the layout (in both languages) under Headline. The most important thing is, of course, to specify which format to use (in the case of the figure, “Tableformat”), and one can specify whether it appears horizontally or vertically (in this case it is type:row, so it will be horizontally). Reference name is used as a mnemonic help for the user, and class is used for fine-tuning the appearance, currently only by DMLF members.

The other two options that can be chosen are somehow more involved, and we will leave them for later.

When one is “navigating” from one entity to another, that means only for base layouts, one can also specify different options (this is the purpose of the choice “Nav. button ordering format”). Currently there is only one option, namely Linking. By choosing it one can browse from one to the next entity using navigation buttons.

A more powerful and imposing capability is the possibility to define sublayouts, using associated entities, as shown in the next layout in Figure 10, where we defined a presentation of works (Obra/Verk) which

should also present the publications of that work we know about, using the association “has been published as” (which in our system is named “publicada como/ har blitt utgitt som”).

Figure 10 A layout with a sublayout

Note that this sublayout also includes a “child layout”, which explains how the publications (and not the works) will be displayed. In addition, it is possible to indicate if the association is to be followed ascending or descending (this is a metaphor for association direction: “ascending” means using the inverse association). It should be obvious that one would have to have defined a layout for publications (and also the corresponding formats) before one could choose it. (One may of course navigate to the Formats bar or to a new Layouts bar if one finds out that they are not yet defined.)

Note also that one can have as many sublayouts as one cares to define. And they can be given a different order by the Order figure which is the leftmost value one can specify. And sublayouts can themselves have sublayouts.

All this will make more sense when we will look at the results of using these layouts and formats later on.

Populating STIG

So far we have only described the skeleton of the STIG instantiation of PANTERA, or better, the model of how it should work. It is time we put some information there so that we can actually search it and

visualize it and supplement it. (To be honest, it really does not make sense to create hundreds of layouts and formats before one sees them in practice, with data. Then one can reshape the visualization.)

There are two ways to input data: the manual one, which is for researchers who are working in a specific part of the whole information system, and are studying a particular translation or work in detail, and will add information along their own research. Or that can be used to edit or add some bits of information here and there.

For that there is a specific mode, called Values, where one can start from scratch and create new entities and add their specific properties.

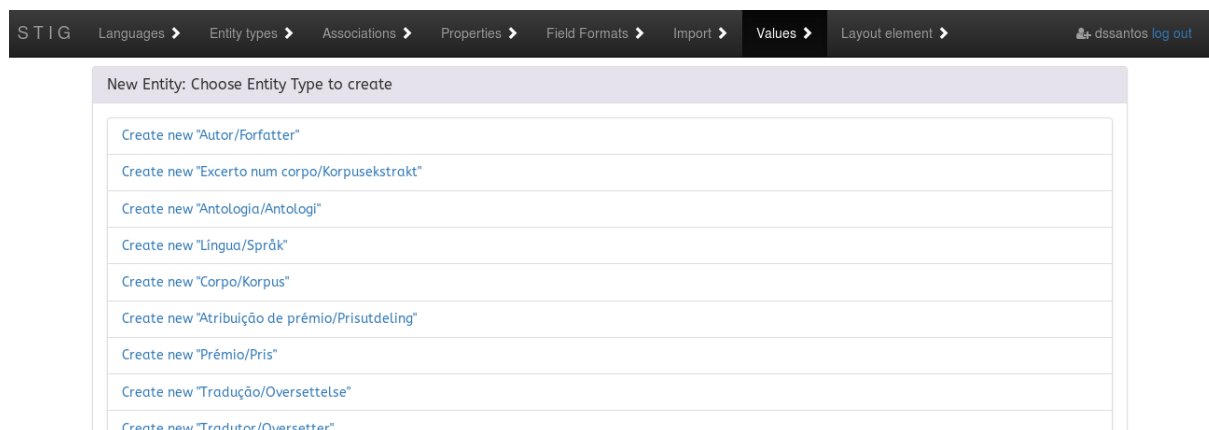


Figure 11 Using the Values interface to add data

Then there is a more powerful form of inputting data from a spreadsheet, or better, from a tab-separated file that is easy to create out of any spreadsheet, where we can specify which columns go to which entities and which properties, and which associations hold among which columns. Often one starts with this kind of tabulated information and proceeds by adding more detailed and specific one later.

This is also what we did for PANTERA, for which we had already a lot of information tabulated. So in this case it is the Import tool that one should use.

Preliminaries to using import

A spreadsheet represents only one kind of entity – in our case a pair of translated works that we use for the PANTERA corpus. In each line we have recorded the publication details for source text and target text included in the corpus. For both publications there is the author name (as represented in the publication), the title, the publisher, the place and the date of publication, variety of language, ISBN, the number of pages, the pages selected for PANTERA. Then there is also the name of the translator(s) and possible intermediary languages. We have also birth dates for authors and their nationality in the lines of the respective works. And for all kinds of values who may differ in Norwegian and in Portuguese (like place name, nationality, genre, etc.) we have two columns. This means that we can in fact obtain information about different kinds of entities from this imperfect knowledge source in this mixed format, provided we can unambiguously indicate which entity types each column refers to, and which associations are implicitly encoded there.

This information is given to STIG by appending to the column title precisely to which entity types and properties and working language it relates, and saving the file in tab-separated UTF8 text. An example of such format can be found in the appendix, but the first nine columns can be detailed here.

```

52__1__68__string__lc__id->korpusekstrakt
    24__1__79__string__lc__navnForfatter1->utgivelse1
24__2__79__string__lc__navnForfatter2->utgivelse2
    1__1__79__string__lc__navnForfatter->forfatter1
    1__1__26__numeric__lc__Nasceu->forfatter1
    46__1__46__string__lc__Título(oversettelse)->utgivelse2
    23__1__29__string__lc__Título(original)->verk
    24__1__38__string__nob__Local(utgivelse1)
24__1__38__string__por__Local(utgivelse1)

```

Translated back to a human format, the first column contains an entity of type 52 (the first one, so 52_1), and the property id (which is number 68), the indication that the value is a string, and that it is common for both language codes (lc stands for language common), and then it is written in a human understandable way to make it easy to check by the developer (or importer). It says: the original name “id” in the spreadsheet relates to the entity type “korpusekstrakt”.

The second column provides the first publication entity type (24), so the 1 after, and the property 79 (name of the author as specified in the publication), which is a string, and does not change with language, so “lc”. It was dubbed “navnForfatter1”, and it is specified that it is a property of “utgivelse1”. (These names are not relevant for STIG, are just to help the understanding the column contents.)

The third column is just like the second, indicating the author name in the published translation. (While in the overwhelming majority of the cases it is the same value, there are cases of difference that made us fill in different slots.)

The fourth column provides a value associated to an author, 1_1, in fact his/her name, property 79, again a string, while the fifth column is the first numeric value and encodes the authors birth date.

The sixth and the seventh columns inform about the translated title and the original title, and are therefore associated to a Tradução/Oversettelse (property title) and Obra/Verk (property title). They are strings, and they are in a particular language, so they do not require two languages, and have therefore lc. The two last columns described here are examples of one value that may have differences depending on the language, namely place of publication. For example, Copenhagen is spelt Kjøbenhavn in Norwegian and Copenhaga in Portuguese from Portugal and Copenhague in Brazilian Portuguese. Or Cristiânia is the Portuguese for Kristiania. So this property of a publication (in this case the first publication, 24_1, has to be encoded in both Norwegian and Portuguese, and therefore the codes nob and por instead of lc.

See Figure 12 to see how to start import:



Figure 12 Using the Import command in STIG

The next step is to import the file itself, as shown in Figures 13, 14 and 15, but making sure that the codes are right:

STIG Languages > Entity types > Associations > Properties > Field Formats > **Import** > Values > Layout element > [dssantos.log out](#)

Import from file:

Entities with belonging variables

Use the below generated labels, on the right-hand side, as the start of headings for your import-file-columns. Add __language-code after the headerstart to indicate which language the column values are in. If the column is the same for both languages, just keep __lc__ instead of any of the languagecodes. **Example:**
 23__30__por__Título

2 Possible language codes(lc): nob | por |

forfatter 1- field/variable

Nome Navn	1__1__1__lc__
Nasceu Født \d(4) \d(4) Contém ano de nascimento, 4 dígitos Inneholder fødselsår, 4-sifra tall	1__1__26__lc__
Morreu Død Ano da morte, 4 dígitos Dødsår, 4 siffer \d(4)	1__1__27__lc__
País de origem Opprinnelsestand Portugal Portugal	1__1__28__lc__

Figure 13 Continuing import

Also associations should be automatically added, therefore they are listed in the import process for the user's convenience:

Associations

Subject type	Association Name	Object type	Association id
Autor/Forfatter (eid:1)	-> escreveu/skrev	Obra/Verk (eid:23)	Id: 15
Obra/Verk (eid:23)	<- cujo autor é/hvis forfatter er	Autor/Forfatter (eid:1)	Id: 15
Autor/Forfatter (eid:1)	-> editou na editora/publiserte i forlag	Editora/Forlag (eid:43)	Id: 29
Editora/Forlag (eid:43)	<- publicou obras de/publiserte verk av	Autor/Forfatter (eid:1)	Id: 29

Figure 14 Listing associations and their ids

And the last job is to specify which columns they relate:

Which association(s) do you have in your import file? Look at the example in the texarea on how to add associations in this import.
 Notation: *subject-type and number__assid__object-type and number;*
 Example: 1__2__15__23__1
 (dvs: forfatter, andre entiteten av 2 (se liste over), assosiasjonsid: 15, skrev, verk, kun en entitet av den typen i denne fila)

1__1__15__23__1, 45__1__19__23__1, 46__1__24__24__2, 23__1__20__24__1, 1__1__18__45__1

Upload file

After making sure you have edited the column headers correctly, and chosen your associations the correct direction, upload your file below.

No file selected.

Figure 15 Specifying which associations to import

Several things should be noted:

- If one imports on top of the contents, one gets duplicate entries
- Is it not possible to undo an import, but it is always possible to ask for deletion of the full contents to the STIG developers, DMLF.

It is however important to stress that you don't create an information system once and for all, this is not the point of STIG at all. Rather, its development was guided by the concern that it should be easy to add new and more complex information to the system, and that several users and researchers should be able to do it, after having agreed on the overarching "information architecture". So after an initial import, one is supposed to edit values in STIG.

We thus proceed to explaining how a normal user (without administration privileges) can interact with a STIG instance. But, as is often the case, users can also be contributors to the

information pool, so it should not be surprising to realize that after searching it is often the case that one wants to change, correct or extend the information obtained. We will come to that later.

How to search in STIG

Contrary to the definition of the entity types, properties and associations, that is something fundamental that has to be decided at an initial stage, and that, therefore, is protected through an administrative layer, search and visualization can (and should) be done by all interested users. (The issue of how to validate user content will be discussed later.)

This is how it was deemed so important to define formats and layouts so that the presentation gives the user what s/he really wants, and not only one rigid table format, as is often the case with other information systems.

Also, we wanted to implement several innovative formats such as time lines and parallel time lines (between translation and source) that make specific kinds of visualization appropriate for the context of translation but not necessarily in other contexts. Not everything is already there, though, as the reader can appreciate in the “Still missing” section in the end of this document.

The entry page(s)

We have to separate the generic system that is called STIG, and which can encompass as many instances as there are interested (super)users, and the use of STIG for a particular domain.

Figure 16 shows the entrance to STIG as a general tool, with the particular instances that exists now.

The screenshot shows the homepage of the STIG system. At the top, there is a black header with the University of Oslo logo. Below this, the main heading reads "STIG System for Translation Information in General". To the right of the heading is a portrait of Stig Johansson. Below the heading is a paragraph of text explaining the system's origin and purpose. The main content area is divided into four project cards: "PANTERA" (with a black panther image), "SEM" (with a small "sem" icon), "Reiser" (with a small "reiser" icon), and "Valen-brev" (with a small "valen" icon). Each card contains a brief description of the project. At the bottom of the page, there is a red circular logo and a footer with the text "Responsible for this service DMLF - HF" and a cookie notice.

Figure 16 First page of STIG

For PANTERA, one can of course click in her box, but normally one would like to go directly there, and this is done by appending the string pantera to the STIG site. So, the address <https://stig.hf.uio.no/pantera> takes us to the entry point to the information about translation between Norwegian and Portuguese. This should be the right practice for other uses of STIG in the future.

The first page a user sees is in Figure 17, and should provide a bird's eyes view of both the kind of information that is stored in the system and an easy way to start browsing it. The slots that are shown are those entity types whose format was defined as “this layout element creates a root menu”, and the information is the one that was stored when we defined the corresponding entity types, and wrote anything in the description slots – see Figure 2.

The screenshot shows the initial page of the STIG system. At the top, there is a black header with the UIO University of Oslo logo on the left and language options 'Norwegian' and 'Portuguese' on the right. Below the header, the title 'STIG' is prominently displayed in large, bold letters, with the subtitle 'System for Translation Information in General' underneath. To the right of the subtitle is an 'Admin' link. Below the title, there are two navigation options: 'Entity Types' (with a dropdown arrow) and 'Advanced Search'. The main content area is divided into four panels, each representing an entity type with a description: 'Autor' (Autors são entidades (pessoas ou organizações) que publicaram obras que foram traduzidas de português para norueguês ou de norueguês para português), 'Tradução' (Uma tradução é uma obra escrita numa língua diferente da original, à qual se pode identificar o original.), 'Tradutor' (Tradutores são todos os que são descritos como autores de uma tradução; responsáveis, portanto, pelo texto traduzido), and 'Obra' (Obras são aqui consideradas produtos intelectuais literários, abstratos, que podem ter várias edições e serem publicados várias vezes.). At the bottom left is the University of Oslo logo. At the bottom right, there is contact information: 'Contact Information', 'E-mail: stig-contact@hf.uio.no', 'Phone: +4722857657', and 'Responsible for this service', 'DMLF - HF'. A dark grey bar at the very bottom contains a cookie notice: 'Application name uses cookies. Find out why, and how you can delete or control cookies' with a small 'x' icon.

Figure 17 The initial page of PANTERA

One should note that one arrives here also when in admin mode, and clicking on STIG (the leftmost link in the administration header).

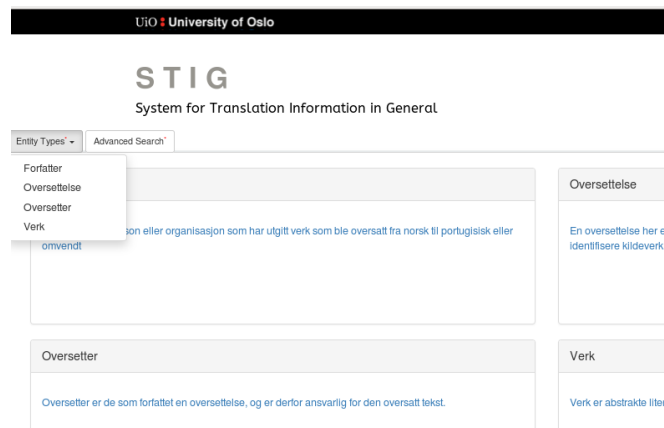


Figure 18 Clicking on the Entity Types tab

If one selects the tab Entity types, one gets to choose also any of these entities (and not others), as can be seen (now in Norwegian mode) in Figure 19. (In addition, there will be in this first page another tab with a general welcome message to give historical information and credits.)

Figure 19 provides an illustration of the result of clicking in Autor in a previous version of PANTERA in STIG.

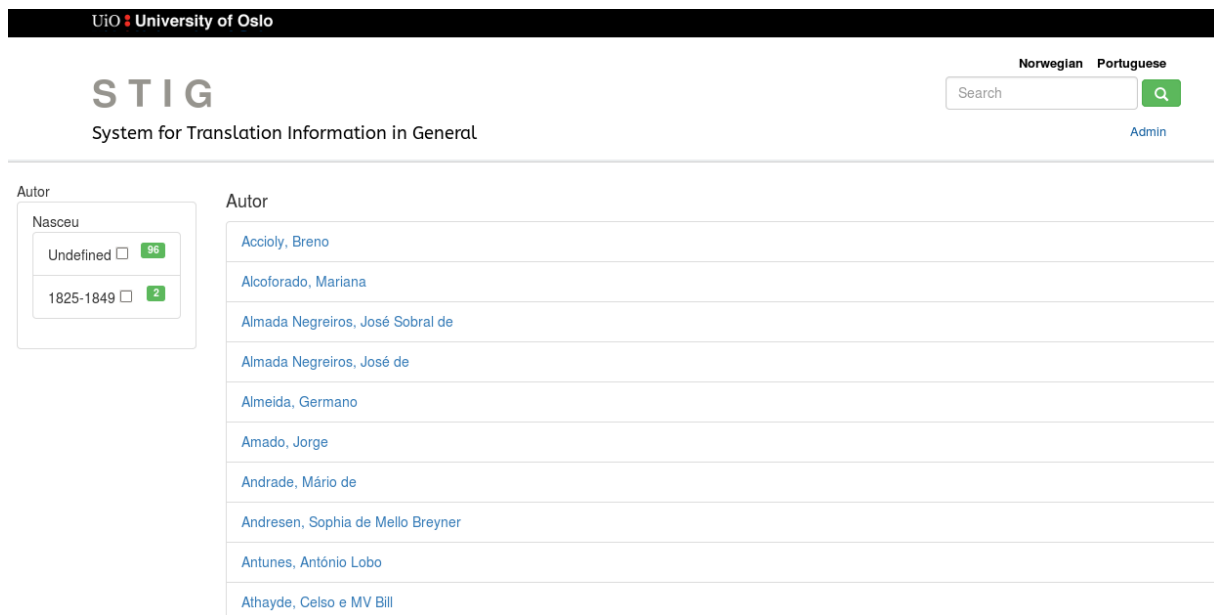


Figure 19 Looking at the list of authors

The reason to show this here is that several problems that are interesting to point out:

First of all, there are two authors which look (and we know are) the same (person), namely the ones called Almada Negreiros. It is not unusual that different publications use slightly different names for an author, and it is even more usual that authors themselves use different pen names for different works. So, one has to decide what the entity type Autor/Forfatter should represent: instances of authors (as many as different strings have been used), or authors with multiple names. In the second case, one of the lines would disappear. In fact, we decided (as might

already been grasped during the description of import above) that we have one author, and different publications may have different strings referring to that author.

Then there is another “author” that clearly are two: Athayde, Celso e (‘and’) M.V. Bill. This is because the column of the spreadsheet really meant authors (most works have a single author), and this had to be corrected inside STIG after the mass import. It also motivated the association co-author between authors, which had not been there from the beginning.

If one clicked then in one the names, one would get the information associated to it, as Figure 20 shows for “Amado, Jorge”. Before analysing the output, it is relevant to point out that they correspond to different layouts. The toothed wheel shows that the values can be edited.

The screenshot shows the STIG interface for the author Amado, Jorge. The header includes the University of Oslo logo and the STIG title. The main content area is divided into sections: 'Autor' (Author) and 'Obras' (Works). The 'Autor' section shows 'Nom: Amado, Jorge' and 'Origem: Brasileiro'. The 'Obras' section lists four works, each with a 'Genre' and 'Titulo' field, and a gear icon for editing. The works listed are: 'Romance' with title 'Capitães da areia', 'Romance' with title 'Dona Flor e seus dois maridos', 'Romance' with title 'Gabriela, cravo e canela', and 'Romance' with title 'Jubiabá'.

Figure 20 After clicking on a particular author

Advanced search

So far we are looking at the data we have, by lists. But the whole purpose of topic maps is to allow for much more complex search, also involving associations, and making use of all sorts of information that may be inside the system in a flexible way.

So Advanced Search is the proof of the pudding, in the sense that it allows as complex searches as the user’s imagination can come up with.

Let us see the initial page of Advanced Search, in Figure 21. And here you can already see that we may look for any kind of entity type, not only the “big four” that were selected for the entry page.

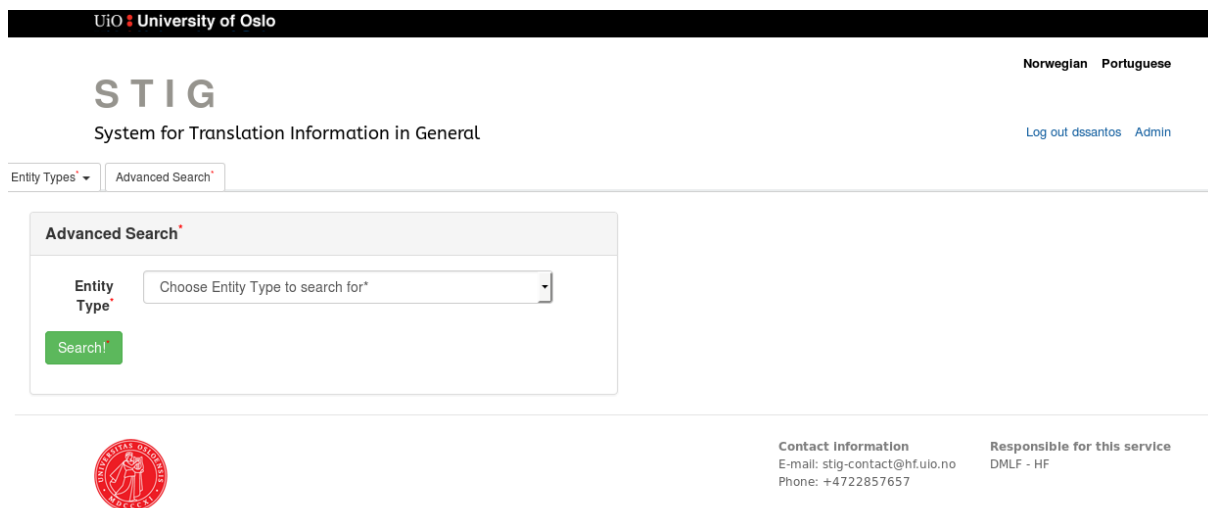


Figure 21 The beginning of Advanced Search

Let us nevertheless choose again Autor/Forfatter, and see that one can specify different values in any of its properties, as shown in Figure 22.

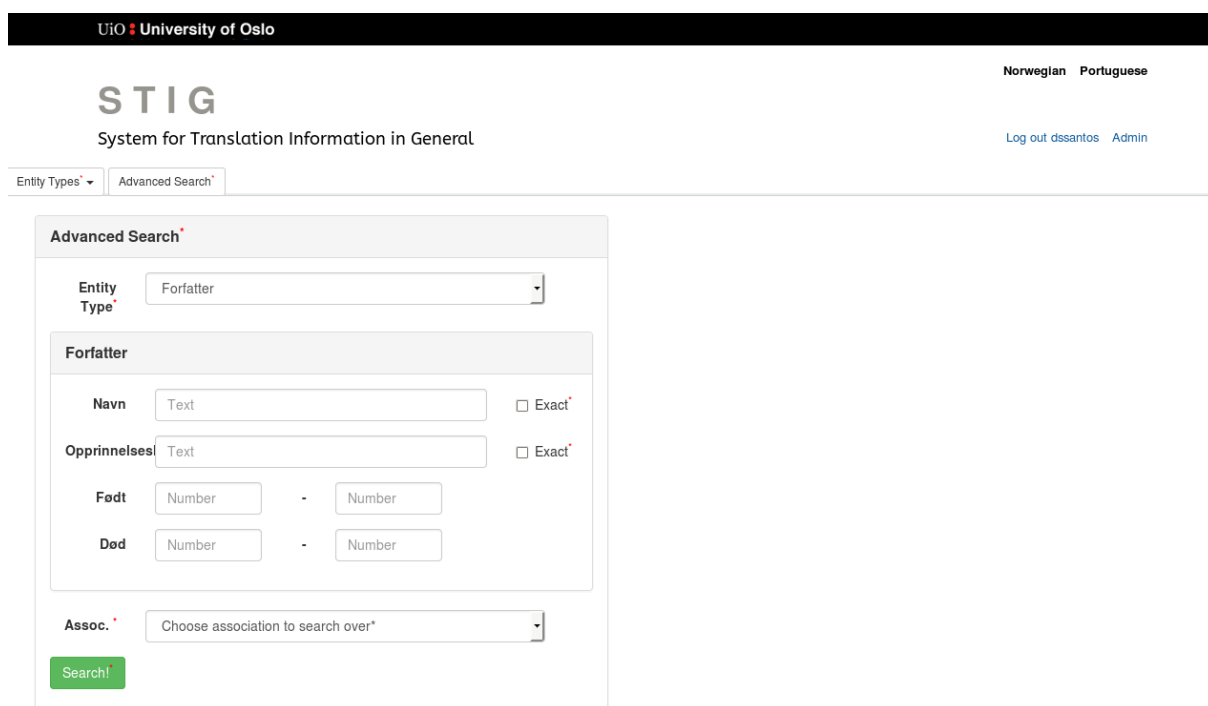


Figure 22 Doing advanced search in authors

One can widen the search to entities related to authors, by choosing one association in the last multiple choice, as the example of Figure 23 illustrates: Brazilian authors who wrote works classified as novels.

And this example can go on, if one should also indicate that those works had been published by a particular publisher, or that they had received a particular prize, or had been translated by a translator born after 1960, and so on.

Entity Types ▾
Advanced Search*

Advanced Search*

Entity Type* Forfatter ▾

Forfatter

Navn Text Exact*

Opprinnelses brasiliansk Exact*

Født Number - Number

Død Number - Number

Assoc.* Skrev Verk ▾

Verk

Tittel Text Exact*

Sted den var skrevet Text Exact*

Periode den var skrevet Number - Number

Genre roman Exact*

Språk Text Exact*

Assoc.* Choose association to search over* ▾

Search!

Figure 23 Advanced search over associations

Before we go to the subject of changing values or creating new entities, let us see the result of such advanced searches – which should be presented in the layout associated with search.

The next Figure, 24, displays the publications with title starting by “O homem”.

STIG

System for Translation Information in General

Log out dssantos Admin

Entity Types ▾ Advanced Search

Advanced Search

Entity Type: Utgivelse

Utgivelse

Dato: -

Tittel: Exact

Kilde eller oversettelse: Exact

Publiserings: Exact

Assoc.: Choose association to search over*

Search Hits

Utgivelse: Forlag ; *

Matching Properties

Tittel: O homem duplicado

Tittel: O homem duplicado

Utgivelse: Forlag ; Lisboa

Matching Properties

Tittel: O homem da guitarra

Tittel: O homem da guitarra

Utgivelse: Forlag ; São Paulo

Matching Properties

Tittel: O homem que amava Yngve

Tittel: O homem que amava Yngve

Figure 24 The result of looking for publications with some string in the title

Changing values and adding new information

The last subject we will discuss at length here is how to change values, and for this we will use the example of the insertion of images (for example book covers), because it also illustrates the power of having properties classified as “Property contains html and display a Rich Text Editor for editing”.

Let us find publications (Publicação/Utgivelse) and define manually their book cover image (Capa/Forside) using plain HTML. Using Advanced Search with Publicação/Utgivelse, and selecting the title we want, we are able to get the right instance. And clicking on the toothed wheel and selecting the property Capa/Forside, we are taken to a rich text editor and add, as in plain Vortex, the URL of the image, as Figure 25 shows.

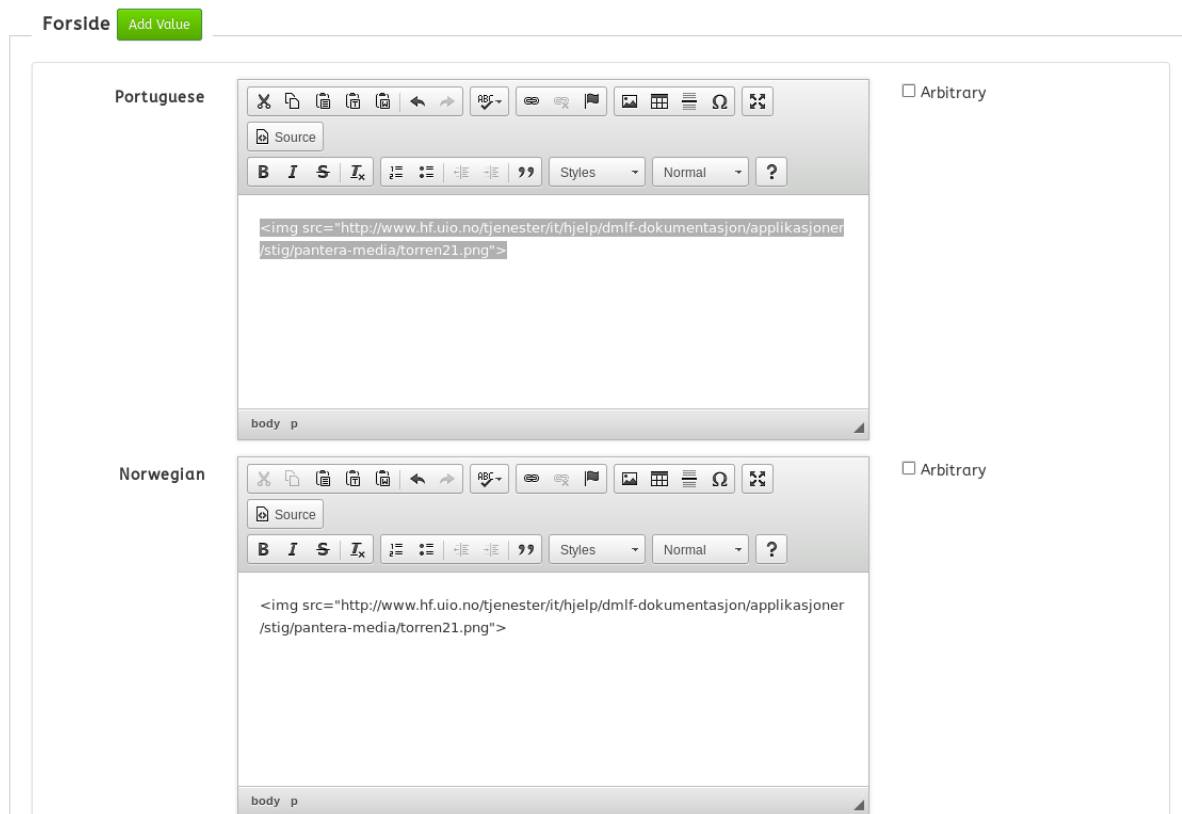


Figure 25 Adding an URL as a value of a property using a rich text editor

Still missing from this text (and from STIG's implementation)

Time lines.

Graphical display of entity types and their associations.

Data export in tab-separated format.

Partial import per entity type.

Source of information for each value if necessary.

Acknowledgements

STIG was developed by Asgeir Nesøen, Damir Nedic and Heidi Løken, to whom I am deeply grateful. It was financed by ILOS, and the data used has been created by several PANTERA research assistants, namely Heidi Jansen, Fernanda Veloso and Peder Østebø.

References

Santos, Diana. "Requirements for PANTERA's information system: Brainstorm at ILOS". Presentation at ILOS, 28 September 2015. <http://www.linguateca.pt/Diana/download/AprReqPANTERA.pdf>

Appendix: a subset of an input file for the import function