Semantic Searching for SignWriting

Steven Aerts (1), Bart Braem (1),
Katrien Van Mulders (2), Kristof De Weerdt (2)
(1) Université d'Anvers
Middelheimlaan 1, B 2020, Anvers, Belgique
{bart.braem, steven.aerts}@ua.ac.be
(2) Université de Gand
Rozier 44, B 9000 Gand, Belgique
{katrien.vanmulders, kristof.deweerdt}@ugent.be

Mots-clefs : analyse de signes, reconnaissance automatique, SignWriting

Keywords: sign analysis, automatic recognition, SignWriting

Résumé Dans cet article nous vous pésentons les résultats de recherches d'un système en ligne de langue des signes indépendant d'annotations manuelles et basé sur SignWriting. La recherche se fait d'une façon intuitives mais flexibles. Les résultats sont classés d'après leur relevance. Le système est en ce moment utilisé pour le dictionnaire de langue des signes flamand contenant plus de 7000 signes.

Abstract In this paper we present the development results of an online sign searching system independent of manual annotations based on SignWriting. Lookup is done on an intuitive yet flexible basis and results are ordered by relevance. The system is currently active for the Flemish Sign Language dictionary containing over 7000 signs.

1 Introduction

We have developed an online database driven dictionary system named `Dixit` currently powering the Flemish Sign Language Dictionary with over 7000 signs(Aerts, Braem, De Weerdt, Van Mulders, 2004-2005; Verweire, 2005). These signs were collected by researchers of the university of Ghent. `Dixit` can convert SignWriting signs typed with SignWriter DOS (Gleaves, Sutton, 1985-2004) to SWML-D. We developed SWML-D as an XML-based representation language for SignWriting dictionaries based on SWML (da Rocha Costa, Dimuro, 2005). The `Dixit` database is modelled on the hierarchical SWML-D structure in order to contain exactly the same information.
The SignWriting system itself is a practical visual writing system for sign languages, composed of a set of intuitive graphical-scadhematic symbols and simple rules for combining them to represent signs(da Rocha Costa, Dimuro, 2005). It was invented by Valerie Sutton inspired by her already created choreographic writing, called DanceWriting (Sutton, 1996-2005; Sutton, 1996-2005). SignWriting symbols represent the body parts involved and the movements and

face-expressions made when producing signs. We based `Dixit` on SignWriting because it is understandable for people who have never seen it before.
In this paper we discuss the internals of our intuitive, user-friendly, yet powerful search by sign system for SignWriting. We started off from the idea that all information is automatically extracted from the signs, without manually enriching them with external information. This machine learning approach overcomes tedious finetuning which is required in other proposed systems (da Rocha Costa, Dimuro, Freitas, 2004).

2 The manual approach

Searching for the meaning of a sign in a database manually enriched with extra semantic information is common practice. It usually consists of selecting the type and direction of the movement, the location on the body where the sign is made and finally the hand form. This information needs to be added to each individual sign (Wilcox Scheibman, Wood, Cokely, Stokoe, 2000), which causes a big slowdown when composing a dictionary with thousands of signs.

3 Semantic view on SignWriting

The first consideration to make is which information can be computed out of SignWriting signs. Movements distinction is reasonable as movements are represented by different symbols. Although trivial for humans determining the moving body part is impossible to implement without an extensive physiological model. The same logic is valid when detecting three-dimensional direction of the movement out of a two-dimensional representation.
When a body part is touched SignWriting gives a very good clue on the zone touched. However when no body part is touched, the location can only be extracted from the symbols by considering the most likely positions. SignWriting depicts a large number of touch variations but most of the time users want more general selection. Thus we allow users to look for the five major SignWriting touch groups (touch, grasp, in-between, strike, brush contact and rub contact).
One very nice property of SignWriting is the accurate and intuitive distinction between hand forms. This is the main feature we search by.

4 Dissection of a search

Everything starts with the user specifying the hand form, which of the five body zones that hand touches (head, torso, arms, legs and hands) and the way in which they are touched, see figure 1. The system can now easily rule out signs that do not contain the selected hand forms, touches if any and body parts. This results in a rather small set of probable signs which can be evaluated thoroughly. This evaluation starts with determining the contact zones and discarding signs without the requested zones. If multiple zones are involved, they have to be matched with the touching body parts.
We have parametrized the semantic goodness of a match to allow ordering by relevance. This goodness-measure is based on the summation of the product of the Euclidean distance between the touch and the middle of the corresponding body zone. We introduced additional improve-

ments based on user feedback: a simple sign will be ranked higher when compared to more complicated signs, very bad matches are dropped to avoid confusion.
The resulting ordering by relevance does not correspond to the SignWriting ordering(Butler, 2001; Sutton, 2004), because for that to be possible information about the dominant hand would be necessary which is impossible to compute without the physiological model. Notice that this search requires no advanced SignWriting knowledge from our users, which turned out positive as it is a publicly available system mostly used as a reference worked by SignWriting novices.

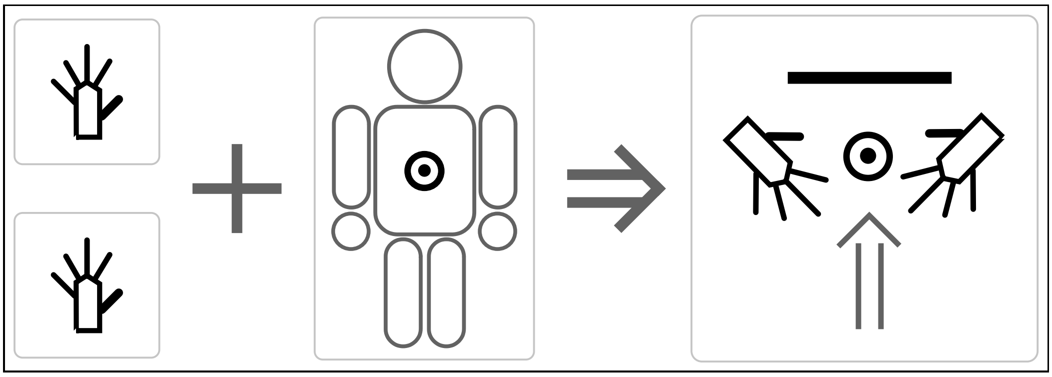

Figure 1: Search for *two opened hands rub torso* returns the Flemish sign for "to enjoy".

5 Performance

Selecting signs with the right symbols and contacts is a database issue where performance is not at stake. Determining the body parts is harder but also be done in the database: in general one thick black horizontal line stands for the chest, whereas two lines can depict the legs and hips.
Calculating the goodness-measure is done over a very limited number of matching symbols and contacts: a sign containing four contacts is extremely rare[1] and will most probably never be searched for by a user. Thus the number of comparisons will be low and will not affect the global performance by one order-of-magnitude. Webserver statistics proved that most queries to take less than 0.5 seconds.

6 Future work

Currently some of the goodness measure parameters are fixed numbers that were finetuned manually while developing. The nature of this problem lends itself perfectly for mapping to the structure of a neural network which would allow an automatic optimization of these parameters (Cohen, Schapire, Singer, 1998). Currently `Dixit` powers the Flemish Sign Language Dictionary however it is highly portable and sign language independent. In the near future we plan to open source `Dixit` to allow wider usage in the Deaf Community, both as a scientific framework for sign language research and as an advanced and intuitive sign language dictionary (Braem, 2000).

[1]We counted 35 signs with 4 or 5 contacts out of 7460 signs.

7 Conclusion

Searching with `Dixit` is intuitive even for a user with very basic SignWriting knowledge. Its friendliness is very high, as proved by user reactions and daily usage. We showed that applying a goodness measure combined with a broadened search makes it a powerful sign language tool. The real strength of the search system lies in the use of the very well specified SignWriting hand forms, which compensates for the vague movements. Because of the use of databases, SWML-D and relatively simple calculations, the method presented allows effecient lookups. Most importantly, the Deaf Community and its researchers will benefit by this new search method since it allows for easier dictionary-searching and linguistic research.

Références

AERTS S., BRAEM B., DE WEERDT K., VAN MULDERS K. (2004-2005), Woordenboek Vlaamse Gebarentaal, `http://gebaren.ugent.be/`.

DA ROCHA COSTA A.C., PEREIRA DIMURO G. (2001), A SignWriting-Based Approach to Sign Language Processing, *The SignWriting Journal*, Vol. 0. 1.

VERWEIRE E. (2005), Vlaamse gebaren gedigitaliseerd (Flemish signs digitalised), *EOS magazine in cooperation with Scientific American*, Vol. 22, 1.

DA ROCHA COSTA A.C., PEREIRA DIMURO G. (2003), SignWriting and SWML: Paving the Way to Sign Language Processing, *TALN 2003 (Batz-sur-mer)*, `http://www10.org/cdrom/posters/p1011/`.

BUTLE C. (2001), An Ordering System for SignWriting, *The SignWriting Journal*, Vol. 0. 1, 1, `http://sw-journal.ucpel.tche.br/number1/article1.htm`.

SUTTON V. (2004), Sutton's SignSpelling Guidelines 2004, *SignWriting Library Database*, 0-914336-83-5.

GLEAVES R., SUTTON V.(1985-2004), SignWriter DOS, `http://www.signwriting.org/`.

SUTTON V. (1996-2005), SignWriting: Read, write, type Sign Languages, `http://www.signwriting.org/`.

SUTTON V.(1996-2005), DanceWriting: Read and Write Dance, `http://www.dancewriting.org/`.

COHEN W.W., SCHAPIRE R.E., SINGER Y. (1998),Learning to Order Things, *Advances in Neural Information Processing System*, Vol. 10.

DA ROCHA COSTA A.C., PEREIRA DIMURO G., BALDEZ DE FREITAS J. (1998), A Sign Matching Technique To Support Searches In Sign Language Texts, *Proceedings of the LREC 2004 workshop for Representation and Processing of Sign Language* .

BOYES BRAEM P. (2000),Sign Language Text Transcriptions and Analyses Using Microsoft Excel, *Center for Sign Language Research, Basel, Switzerland* .

WILCOX S., SCHEIBMAN J., WOOD D., COKELY D., STOKOE W.C. (2000),Multimedia dictionary of American Sign Language, *Proceedings of the first annual ACM conference on Assistive technologies table of contents* 0-89791-649-2.