



Llywodraeth Cymru
Welsh Government



PRIYSGOL
BANGOR
UNIVERSITY

Best Practices for Sharing Language Technology Resources in Minority Language Environments

Delyth Prys

Bangor University, Wales

d.prys@bangor.ac.uk

The Language Technologies Unit, Bangor University

- Self-funded research unit
- Established as the Centre for the Standardization of Welsh Terminology in 1993
- Expanded scope from 2001 onwards to include Speech and Language Technologies
- Focus on Welsh and other minoritized/less-resourced languages

The Welsh National Terminology Portal

<http://techiaith.org/>

The screenshot shows the homepage of the portal. At the top left is a large white 'C' logo. To its right, the text 'WELSH NATIONAL LANGUAGE TECHNOLOGIES PORTAL' is displayed. On the far right, there are logos for 'CYMRAEG' (Welsh) and 'PRIFYSGOL BANGOR UNIVERSITY'. Below the header is a navigation bar with links: HOME, CORPORA, TRANSLATION, SPEECH, API SERVICES, WIDGETS, GITHUB, BLOG, CONFERENCE, and ABOUT. A large photograph of students working on laptops is the main visual element. Overlaid on the photo is the text 'What is the Language Technologies Portal?' and a description: 'The Language Technologies Unit's resources available easily, freely and openly for learning, hacking and commercial activities.' At the bottom of the page, there are three sections with teal headers: 'WHAT KIND OF RESOURCES?', 'FOR WHO?', and 'USING THE RESOURCES'.

CYMRAEG

PRIFYSGOL BANGOR UNIVERSITY

HOME CORPORA TRANSLATION SPEECH API SERVICES WIDGETS GITHUB BLOG CONFERENCE ABOUT

What is the Language Technologies Portal?

The Language Technologies Unit's resources available easily, freely and openly for learning, hacking and commercial activities.

WHAT KIND OF RESOURCES?

FOR WHO?

USING THE RESOURCES

WHAT KIND OF RESOURCES?

The Language Technologies Portal consists of a wide range of resources for the Welsh language, including:

FOR WHO?

The resources are useful to a number of projects and Welsh language software developed by :

USING THE RESOURCES

The resources are open and freely available

The Problem

- Lack of easily accessible tools for developing LT for Welsh
- Expense of developing tools for languages which are not commercially interesting
- Many useful LT tools/resources developed over the years for internal use not available to others
- Lack of knowledge by software/web developers etc on how to use these resources

The Solution

- Hold an audit of all tools/resources developed by the LTU that are available for reuse
- Select the most important ones for repackaging and dissemination
- Create a free, accessible on-line portal
- Include documentation, tutorials and ideas for use

Initial resources selected

1. Social web corpora (Twitter and Facebook)
2. Welsh<>English Machine Translation
3. Welsh<>English Aligner for preparing parallel texts
4. Welsh Text-to-Speech API service
5. Welsh Festival Speech Synthesis resources
6. Spelling and grammar checker API and widget
7. PoS Tagger API
8. Language Detection API
9. ‘Vocab’ bilingual dictionaries widget



WELSH NATIONAL LANGUAGE TECHNOLOGIES PORTAL



[CORPORA](#) [TRANSLATION](#) [SPEECH](#) [API SERVICES](#) [WIDGETS](#) [GITHUB](#) [BLOG](#) [CONFERENCE](#) [ABOUT](#)

TRANSLATION

Language technologies are currently revolutionizing the world of translation. Here at the Language Technologies Unit, we believe that Welsh translators need to be able to understand and use these disruptive new technologies effectively.

We believe that the industry should use translation technologies on the following basis:

- Domain-specific translation engines
- Data from open and closed sources
- Ownership and sharing
- Standards, research and training
- Prioritizing the needs of Welsh translation customers

The Language Technologies Portal contains many free and open resources that can contribute towards achieving these aims. We hope that the Portal will lead to the growth and development of a community of developers and practitioners within the Welsh translation industry.

WHAT RESOURCES ARE AVAILABLE?

These resources for facilitating translation are available from the Language Technologies Portal:

- [Welsh <>> English Machine Translation with Moses-SMT](#)
- An [aligner](#) for preparing data for training Moses-SMT.

Moses-SMT Welsh <> English Machine Translation

This project contains the code, scripts and related documentation which facilitates the use of Moses-SMT for Welsh.

The Project Structure

- **docs** - contains documentation on how to use the scripts. Specifically, how to:
 - install Moses-SMT from docker.com and download the Language Technologies Unit's translation packages [click here...](#)
 - install Moses-SMT on a normal computer - [click here...](#)
 - run Moses-SMT on a normal computer - [click here...](#)
 - train Moses-SMT and create your own engines - [click here...](#)
- **get** - contains scripts that are necessary to prepare and install Moses-SMT on a Linux computer
- **mtdk** - contains the scripts that support Moses-SMT's training features

The Project Files

- **Dockerfile** - the file that is used to create docker images



Search GitHub

Explore Features Enterprise Blog

Sign up

Sign in



Porth Technolegau Iaith Cenedlaethol Cymru

Casgliad o adnoddau agored ac am ddim technolegau iaith Cymraeg ar gyfer clybiau codio, hacwyr, a chwmniau i greu cynnyrch digidol Cymraeg

📍 Uned Technolegau Iaith, ...

🔗 http://www.techiaith.org

✉️ techiaith@bangor.ac.uk

Filters ▾

Find a repository...

People

0 ➔

This organization has no public members.
You must be a member to see who's a part
of this organization.

moses-smt

Python ★ 1 ⚡ 1

Adnoddau Cyfleithu Peirianyddol Cymraeg / Machine Translation for Welsh

Updated 9 days ago

alinio

Python ★ 0 ⚡ 0

Cod hwyluso alinio gyda hunalign a dogfennaeth ar sut i ddefnyddio LFAAligner / Code
for simplifying using hunalign and documentation for LFAAligner

Updated 28 days ago

festival

HTML ★ 0 ⚡ 0

Dogfennaeth ac enghreifftiau ar sut i ddefnyddio API testun i leferydd Cymraeg trwy
Festival ~~ Documentation and examples on how to use the Welsh text-to-speech API

Updated 28 days ago

llais_festival

Scheme ★ 1 ⚡ 0

Llais Deuffon Festival Cymraeg - Welsh Diphone Festival Voice

Updated 28 days ago

PorthTechnolegaulaith/turing-test-lesso... + GitHub, Inc. (US) https://github.com/PorthTechnolegaulaith/turing-test-lessons Google

GitHub

This repository Search

Sign up Sign in

PorthTechnolegaulaith / turing-test-lessons

forked from raspberrypilearning/turing-test-lessons

Watch 3 Star 0 Fork 10

Mae'r cynllun gwaith hwn wedi'i gynllunio fel cyflwyniad i raglennu seiliedig ar destun gan ddefnyddio Python ar y Raspberry Pi. Erbyn diwedd yr holl wersi, bydd myfyrwyr wedi creu robot sgwrsio rhwngweithiol ac wedi'u cyflwyno at technolegau iaith Cymraeg yn ogystal a'r cysyniad cyfrifiadurol o ddeallusrwydd artifisiai / This scheme of work ha...

2 commits 2 branches 0 releases 2 contributors

branch: master turing-test-lessons / +

This branch is 2 commits ahead, 73 commits behind raspberrypilearning:master

Ychwanegu 'technolegau iaith Cymraeg' i'r README

techiaith authored on Mar 26 latest commit c83074df38

Lesson-1	Cyfeithiad llawn / Completed translation	a month ago
Lesson-2	Cyfeithiad llawn / Completed translation	a month ago
Lesson-3	Cyfeithiad llawn / Completed translation	a month ago
Lesson-Garn	Cyfeithiad llawn / Completed translation	a month ago

Code Pull requests Pulse Graphs

HTTPS clone URL <https://github.com/PorthTechnolegaulaith/turing-test-lessons>

You can clone with HTTPS or Subversion.

Clone in Desktop Download ZIP



'Vocab' widget installed on newspaper website

Cynyddu maint testun

golwg360

RSS • VOCAB 

 Swyddfa Ewrop
Creadigol
y Deyrnas Unedig
& o'mr
www.creativeeuropeuk.eu

 Creative
Europe
Desk UK
Wales

EWROP GREADIGOL
Cyllid a chyfleoedd i'r sectorau
diwylliannol, creadigol a chlyweledol

 Aminiaeth y Bwrdd
Llywodraeth Cymru
Part Funded by
Welsh Government

Newyddion Chwaraeon Celfyddydau Calendr Blog Bwyd Lle Pawb Lle Busnes Adran Swyddi

Cymru | Prydain | Rhyngwladol

Addysg | Arian a Busnes | Gwleidyddiaeth | Iechyd | Rhyfeddodau | Gwasanaethau Brys | Cefn Gwlad | Gwyddoniaeth a Thechnoleg | Iaith | Clonc360

Cwmni cyfieithu yn agor yn Aberystwyth

Ebrill 24, 2015 09:50

Mae un o brif gwmniau cyfieithu'r gorllewin yn cynnal lansiad i ddathlu agor swyddfa newydd a chreu swyddi newydd yn Aberystwyth heddiw.

Dros y tair blynedd ddiwetha' mae cwmni Trywydd wedi gweld "cynnydd yn y galw" am ei wasanaethau yn ôl y rheolwyr, ac mae hynny wedi arwain at greu pedair swydd newydd.

Fe fydd yr Aelod Cynulliad lleol Elin Jones yn bresennol yn y lansiad sy'n cychwyn am 11.

Mae wyth o bobol yn gweithio yn swyddfeydd y cwmni yng Nghaerfyrddin a Llandeilo.

Cyflogaeth leol

Ffrwyth prosiect Ewropeaidd o'r enw Dewis, a sefydlwyd yn 2002, yw Trywydd.

Dyweddodd y Prif Weithredwr Owain Gruffydd: "Rydym yn falch iawn ein bod fel cwmni yn ehangu ymhellach, yn dilyn y galw cynyddol am ein gwasanaethau cyfieithu a chynllunio iaith – nid yn unig yn yr ardal hon, ond ledled Cymru.



Straeon Perthnasol

Cwmni iogwrt yn wynebu colli swyddi

Aberystwyth yn brifddinas diwylliant Prydain?

Tridiau o streic yn Ysgol Penweddig

Trigolion Wrecsam yn yfed mwy o de nag unrhyw le arall ym Mrhydian

Chwilio Golwg 360...



 y ganolfan cynllunio iaith
welsh centre for language planning

cyngor

ymchwil

hyfforddiant

digwyddiadau

prosiectau

+44 (0) 1239 711668
post@iaith.eu www.iaith.eu



'Vocab' dictionary selected on text

CyfeithuCymru Porth Ternau | PORTH TE... Cwmni cyfieithu yn agor y... +

www.golwg360.com/newyddion/cymru/184440-cwmni-cyfieithu-yn-agor-yn-aberystwyth?main_testun=

Rhannu Lleihau maint testun RSS VOCAB

golwg360

Sorbyddiau Europ
Graddig
Cymru
Creative Europe
Cymru
www.creativeeuropeuk.eu

EWROP GREADIGOL
Cyflwr a chyflwr i'r sektorau
diwylliantol, credydol a chyweliadol

Mae'n un o nodweddion y
Perchnogion
Mae'n un o nodweddion y
Perchnogion
Mae'n un o nodweddion y
Perchnogion

Newyddion Chwaraeon Celfyddydau Calendr Blog Bwyd Lie Pawb Lie Busnes Adran Swyddi

Cymru | Prydain | Rhngladol

Addysg | Arian a Busnes | Gwleidyddiaeth | Iechyd | Rhyfeddodau | Gwasanaethau Brys | Cefn Gwlad

Gwyddoniaeth a Thechnoleg | Iaith | Clonc360

Cwmni cyfieithu yn agor yn Aberystwyth

Ebrill 24, 2015 09:50

Mae un o brif gwmniau cyfieithu'r gorllewin yn cynnal lansiad i ddathlu agor swyddfa newydd a chreu swyddi newydd yn Aberystwyth heddiw.

Dros y tair blynedd ddiwetha' mae cwmni Trywydd wedi gweld blynedd eb year, wasanaetha'r Geiriau Cyffredinol Cysgair arwain at grwpiau. Chwilio pellach am blynedd o fewn y Porth Ternau.

Fe fydd yr Aelod Cynulliad lleol Elin Jones yn bresennol yn y lansiad sy'n cychwyn am 11.

Mae wyth o bobol yn gweithio yn swyddfeydd y cwmni yng Nghaerfyrddin a Llandeilo.

Cyflogaeth leol

Ffrwyth prosiect Ewropeaidd o'r enw Dewis, a sefydlwyd yn 2002, yw Trywydd.

Chwilio Golwg 360..

Chwilio Golwg 360..

Coleg Cymraeg Cenedlaethol

Mynd i Brifysgol yng Nghymru fis Medi?

Rho gynrig ar Ysgolion a
Cymhellant y Coleg Cymraeg
o £1,500 cyn 8 Mai 2015

www.sainwales.com DVD NEWYDD

NYTH CACWN

EN 12:13 24/04/2015

Conclusions

- Successfully addresses needs of both local and international developers
 - Hence National Portal + use Github and other international repositories
 - APIs and widgets have proved very popular (due to demand from web developers?)
- Increases interest in open source LTs such as Moses and Festival
- Further queries from public organisations show the need to prioritize ease of use to end users