# 

**Cyfanswm amser**

30 to 45 Munud

**Cynulleidfa**

3 - 100 o bobl. Oedrannau 12+.  
*Wedi ei ddylunio am raddau 6 - 12, dosbarthiadau Addysg Uwch, Cyrff Newyddion, Cyrff Di-elw, a Gweithdai Cymunedol. Does dim angen unrhyw brofiad blaenorol gyda data.*

**Gofod**

* Taflunydd a chyfrifiadur.
* Gallu i ymrannu’n grwpiau bach o 3 o amgylch cyfrifiadur.
* Byrddau mawr neu lawr, neu dâp i lynu papur wrth waliau fel y gall cyfranogwyr ddarlunio

**Cyflenwadau**

* Cyfrifiaduron

*1 am bob 3 chyfranogwr*

* Darnau mawr o bapur

*2 droedfedd x 3 droedfedd yn fras*

* Crayonau trwchus neu farcwyr

**Beth yw SameDiff?**

Mae SameDiff yn cymharu dau neu fwy o ffeiliau testun ac yn dweud wrthych chi ba mor debyg neu wahanol maen nhw. Mae’n eich helpu i weld gwahaniaethau a thebygrwyddau yn y geiriau sy’n cael eu defnyddio ym mhob ffeil fel y gallwch ddysgu am ddadansoddiad meintiol testun. Mae’r gweithgaredd ymarferol yma’n helpu cyfranogwyr i fagu eu llythrennedd data drwy gymharu geiriau dau gerddor a dyfeisio cân newydd byddan nhw’n ei chyd-ysgrifennu.

**Nodau Dysgu**

* Mwy o allu i ddadansoddi data testun.
* Deall bod cymharu dau beth yn ffordd rymus o ganfod storïau mewn data.
* Ymwybyddiaeth o’r math o gwestiynau gallwch chi/dylech chi eu gofyn i ddata testun.
* Deall y gall dadansoddiad algorithmig ddatgelu gwybodaeth ddiddorol am eich data.

**Ysgrifennu’ch Cân Eich Hun *Defnyddio SameDiff i ddadansoddi arddulliau geiriol dau***

***gerddor a llunio deuawd amdanyn nhw***

**Cynnal y Gweithgaredd**

Datrys Problem  
Mae dadansoddi testun mawr yn anodd i’w wneud â llaw. Un ffordd o ddeall “corpws” o destun yw ei gymharu ag un arall, neu i gymharu rhannau ohono. Mae cyfrifiadurwyr wedi dyfeisio dulliau i helpu, gan greu amrywiol ryseitiau, neu “algorithmau”, sy’n gallu cymharu dau gorpws. Mae SameDiff yn rhedeg rhai o’r algorithmau hynny i chi fel y gallwch chi geisio cymharu dau ddarn mawr o destun â’i gilydd.

Rhannu Enghreifftiau i Ysbrydoli

Mae data testun mawr ym mhobman o’n cwmpas. Heddiw gallwch chi lawrlwytho holl e-byst Ysgrifennydd Gwladol Hillary Clinton, ceblau diplomataidd gan Wikileaks, neu holl nofelau Sherlock Holmes o Brosiect Gutenberg. Mae dadansoddi a delweddu’r testunau mawr yma’n beth cyffredin i’w wneud erbyn hyn, mewn ffyrdd difrifol neu ddifyr. Dangoswch Jaz Parkinson’s “Color Signatures” sy’n cymharu’r lliwiau a enwir mewn llyfrau gwahanol ([*http://jazparkinson.tumblr.com*](http://jazparkinson.tumblr.com)) (Saesneg yn unig), a Tahir Hemphill’s “Rap Research Lab” ([*http://rapresearchlab.com*](http://rapresearchlab.com)) (Saesneg yn unig).

Lansio’r Gweithgaredd

**I’ch atgoffa**

* Rydym ni’n rhedeg algorithmau bob dydd. E.e. pan gollwch chi’ch allweddi rydych chi’n rhedeg algorithm i chwilio amdanyn nhw – gwirio’ch pocedi yn gyntaf, y bwrdd wrth ymyl y drws, ac ati
* Mae tebygrwydd cysein o 1.0 yn golygu yn union yr un peth; mae sero yn golygu cwbl wahanol.

**Termau i’w Cyflwyno**

**Algorithm**

*Set o gamau i chi (neu gyfrifiadur) eu cymryd i ddatrys problem.*

**Corpws***Casgliad o destunau ysgrifenedig. Er enghraifft, yr holl eiriau yng nghaneuon Katy Perry.***Tebygrwydd Cysein***Mae sgôr Tebygrwydd Cysein yn ceisio dweud wrthych chi pa mor debyg mae dwy ddogfen ar sail nifer y gweithiau mae geiriau’n cael eu defnyddio yn y ddwy.*

**Cynnal y Gweithgaredd (parhau)**

Cyflwyno’r Offeryn  
Agorwch SameDiff ([*https://databasic.io/samediff*](https://databasic.io/samediff)) a dewiswch Yws Gwynedd ac Dafydd Iwan o’r samplau. Ar y dudalen ganlyniadau esbonwich fod y golofn chwith yn dangos geiriau sy’n unigryw i Yws Gwynedd, a’r golofn dde yn dangos geiriau sy’n unigryw i Dafydd Iwan. Dyna eu gwahaniaethau. Mae’r golofn ganol yn dangos y geiriau sydd ganddyn nhw’n gyffredin. Tynnwch eu sylw at frig y dudalen ganlyniadau lle mae’n dweud, “Mae'r ddwy ddogfen yma yn eithaf tebyg”. Mae SameDiff yn defnyddio algorithm o’r enw “tebygrwydd cysein” i roi sgôr tebygrwydd i chi. Mae tebygrwydd cysein yn gweithio drwy greu rhestr o eiriau gan Yws Gwynedd a rhestr o eiriau gan Dafydd Iwan. Mae’n cyfrif pa mor aml mae pob term yn ymddangos ym mhob dogfen ac yna’n cymharu ba mor agos mae’r ddwy restr yn cyd-fynd a’i gilydd. Mae’n algorithm defnyddiol i ddadansoddi testunau.

1. Mae gan gyfranogwyr 15 munud.
2. Mae cyfranogwyr yn gweithio mewn timau o dri.
3. Mae pob tîm yn defnyddio SameDiff i gymharu geiriau caneuon dau gerddor. Gan fod cydweithrediadau cerddorol yn boblogaidd iawn, dewiswch ddau artist a dychmygwch sut gân fyddai’r ddau yn ei hysgrifennu gyda’i gilydd. [h*ttps://databasic.io/samediff*](https://databasic.io/samediff)

Brasluniau sampl



1. Mae pob tîm yn yn ysgrifennu geiriau eu cân ar ddarn mawr o bapur gyda chrayonau.
2. Mae timau’n ennill pwyntiau bonws os: (a) mae eu cân yn odli a/neu (b) maen nhw’n dyfeisio tôn i ganu’r gân iddi a/neu (c) maen nhw’n ei pherfformio yn null karaoke ar gyfer y grŵp.

Rhannu Adborth  
Cymerwch 1 funud i bob grŵp rannu eu cân newydd. Rhai cwestiynau a themâu i chwilio amdanyn nhw a chanolbwyntio arnyn nhw yn ystod y drafodaeth:

* Wnaethoch chi sylwi ar unrhyw themâu cyffredin?
* Ydy’r geiriau canlyniadol yn fwy diddorol pan maen nhw’n dod o artistiaid â gwaith sy’n wahanol iawn?
* Mae cymharu yn ffordd rymus o ganfod storïau mewn data.
* Gall gweithio gyda data fod yn hwyl!