

Golygu Digidol: Profiad Prosiect Guto'r Glyn

Ann Parry Owen

(Ymddangosodd yn *Tu Chwith*, Gwanwyn 2013)

Mae ysgolheigion wedi gweld potensial dulliau digidol o gofnodi, cyfnewid a chyhoeddi hen destunau ers dyddiau cynharaf cyfrifiadura yn yr 1940au. Prosesu rhifau oedd pwrpas y cyfrifiaduron cynharaf: rhywbeth eilradd oedd y testun ei hun, ac un o'r prif broblemau oedd sut i ddynodi'r amrediad eang o lythrennau a symbolau. Roedd y setiau nodau cyntaf yn gyfyngedig i briflythrennau Saesneg America, heb arwyddion fel '£' hyd yn oed, heb sôn am lythrennau acennog. Serch hynny aeth nifer o ysgolheigion arloesol ati i ddyfeisio ffyrdd i baratoi testunau fel y gellid eu mynegeo'n gyfrifiadurol. Daeth cyfnewid testunau'n broblem gynyddol wrth i unigolion ddyfeisio eu systemau unigryw eu hunain i'w paratoi. Dyblygwyd gwaith yn aml gan ei bod yn haws paratoi testun o'r newydd yn hytrach na cheisio deall ac addasu system rhywun arall. Dyfeisiwyd nifer o systemau amgodio ('encoding') gwahanol ar gyfer prosiectau mewn gwahanol ganolfannau, a gwelwyd gwir angen am un system safonol, y byddai pawb yn ei dilyn er mwyn gallu cyfnewid testunau ac elwa ar waith ei gilydd.

O'r diwedd, tua 1987, dechreuwyd cydnabod un o'r systemau amgodio hyn yn safon, sef y Text Encoding Initiative (TEI),¹ system a oedd yn seiliedig ar safon ryngwladol ISO SGML ('Standard Generalized Markup Language').² System o safonau i greu ieithoedd 'markup' yw SGML, a'r iaith fwyaf adnabyddus sy'n seiliedig arni yw HTML ('Hypertext Markup Language') sydd wedi bod yn sylfaen i'r We Fyd Eang o'r 1990au hyd heddiw. Tua'r un adeg daeth Unicode yn safon ryngwladol ISO a fabwysiadwyd gan bob gwneuthurwr cyfrifiaduron o bwys.³ Am y tro cyntaf roedd safon ar gyfer dynodi holl lythrennau acennog y Gymraeg. Nod Unicode yw cynrychioli pob system ysgrifennu gyfoes yn y byd, gyda lle i dros 65,000 o symbolau gwahanol. Yn ddiweddar ychwanegwyd llawer o symbolau ar gyfer trawsysgrifio testunau'r Oesoedd Canol, gan gynnwys y llythyren sy'n edrych fel '6' agored (U+1EFD) sy'n gallu cynrychioli 'u', 'v' neu 'w' mewn hen destunau Cymraeg.

Cyhoeddwyd XML yn 1998 (ar sail SGML eto) i hwyluso cynrychioli data ar y We, ac mae wedi lledu'n gyflym i bob math o feysydd cyfrifiadurol. Gwelodd datblygwyr y TEI fanteision XML dros SGML, ac yn 2002 newidiwyd canllawiau'r TEI i gwmpasu XML ac Unicode. Prif gynnyrch y TEI yw set o ganllawiau ar gyfer mathau gwahanol o destunau: rhyddiaith, barddoniaeth, drama, geiriaduron, ac yn y blaen. Trwy amgodio testun yn dilyn y canllawiau, gellir sicrhau y bydd pobl eraill yn gallu ei ddeall a'i ddefnyddio yn y dyfodol. Cyhoeddwyd y canllawiau cyntaf (P3) yn 1994.

Erbyn hyn mae cyrff cyllido ymchwil yn sylweddoli manteision defnyddio'r TEI i ddiogelu eu buddsoddiad mewn ymchwil, ac yn argymhell (os nad yn mynnu) bod ymchwilwyr yn defnyddio'r system i baratoi testunau digidol. Felly pan luniwyd cais chwe mlynedd yn ôl i Gyngor Ymchwil y Celfyddydau a'r Dyniaethau (AHRC) gan Ganolfan Uwchefrydiau Cymreig a Cheltaidd Prifysgol Cymru ar gyfer cynnal tîm ymchwil i baratoi golygiad digidol newydd a dwyieithog o farddoniaeth Guto'r Glyn (c.1435–c.1490) ar gyfer y We, fe wnaethom ninnau ymrwymo i amgodio'r testun gan ddilyn canllawiau'r TEI.

Un o wendidau'r TEI (er y byddai rhai yn ystyried hyn yn gryfder) yw nad oes canllawiau ar ba feddalwedd i'w defnyddio i baratoi'r testunau yn y lle cyntaf. Er bod XML yn boblogaidd iawn ym myd busnes a thu hwnt, mae prinder offer hylaw ar gyfer trin ein math ni o ddata. Tasg gymharol syml ond llafurus yw amgodio testun yn ôl canllawiau'r TEI, ond

mater arall yw addasu'r data pe bai angen ei olygu. Ceir nifer o raglenni soffistigedig iawn ar gyfer rhaglenwyr a datblygwyr gwefannau, ond maent yn llawer mwy cymhleth na rhaglenni swyddfa fel Microsoft Word, ac fel tîm ymchwil, roeddem ni'n awyddus i fuddsoddi'r rhan fwyaf o'n hamser prin i ddehongli barddoniaeth Guto. Felly penderfynwyd paratoi'r data yn y dull traddodiadol gyda Word, a'i gadw yn y ffurf honno nes iddo fod trwy'r holl gamau golygyddol. Unwaith roedd y testun yn weddol barod, cadwyd un set o ffeiliau Word yn sail i'r cyfrolau print arfaethedig, a set arall i'w hamgodio ar gyfer y We. Yn ddamcaniaethol, dylai fod yn bosibl i gynhyrchu testun XML a'i ddefnyddio yn sail i'r fersiwn print hefyd; ond, oherwydd problemau cyfieithu'r XML yn ôl i destun wedi ei fformatio ar gyfer ei argraffu, a hefyd oherwydd natur wahanol y ddau gynnyrch, roedd yn haws trin y ffeiliau XML a'r ffeiliau Word ochr yn ochr o'r pwynt hwn ymlaen, er bod hynny'n golygu bod angen newid dwy set o ffeiliau os oedd angen cywiro neu addasu'r testun. Ar gyfer golygu'r XML rydym wedi dilyn un o argymhellion golygyddion y TEI a defnyddio'r feddalwedd oXygen gan SyncRO Soft Ltd. o Craiova yn Romania,⁴ rhaglen a ddefnyddir mewn llawer o brifysgolion ledled y byd. Fe wnaethom gyflogi Dr Jenny Day yn Olygydd Amgodio i ymgymryd â'r gwaith hwn, ac i gydweithio'n glòs gyda'n Swyddog Gwe, Alexander Roberts o Brifysgol Abertawe sydd wedi creu strwythur y wefan i ni. Mr Roberts fu hefyd yn gyfrifol am greu Gwefan Dafydd ap Gwilym (<http://www.dafyddapgwilym.net>). Rhaglen Anastasia a ddefnyddiodd yn sylfaen i'r wefan honno, rhaglen a oedd yn bwerus ac arloesol iawn yn ei hamser pan ddatblygwyd hi gan Peter Robinson yn 1993 ar gyfer golygu gwaith Chaucer.⁵ Ond er mwyn gwneud Gwefan Guto'r Glyn yn fwy hygyrch nawr ac yn y dyfodol, defnyddiodd Mr Roberts feddalwedd ddiweddarach gan gynnwys fframwaith Django, technoleg Oracle Berkeley DB XML a'r iaith raglennu Python. Seiliodd y rhyngwyneb ar CSS 'Blueprint' gan Joshua Clayton a fframwaith JavaScript JQuery UI. Er mwyn hyrwyddo mynediad i'r testun ar gyfer ymchwilwyr, datblygwyd API ('Application Programming Interface') sy'n caniatáu i ddeunydd gael ei dynnu'n uniongyrchol o'r wefan i wefan arall (fel a wneir yn achos gwefan 'Cymru Guto' yr ydym ni yn ei datblygu ochr yn ochr â gwefan cerddi Guto am ddiwylliant materol Cymru ar ddiwedd yr Oesoedd Canol), neu i greu dulliau newydd o gyflwyno neu o weithio gyda'r data (fel sail, er enghraifft, i adnoddau neu 'Apps' ar gyfer teclynnau mudol). Prosiect pum mlynedd fu hwn ac ym mis Medi 2012, lansiodd Gwefan Guto'r Glyn (<http://www.gutorglyn.net>): fersiwn beta ydyw ar hyn o bryd, gan ein bod yn dal i ychwanegu deunydd a mireinio'r wefan wrth ymateb i awgrymiadau defnyddwyr.

Mae gan y cyfrifiadur, felly, lawer i'w gynnig i'r golygydd testunau: y gallu i chwilio, i **drefnu**, i greu mynegeion geirfaol, i dynnu ar adnoddau digidol eraill, i gynnwys delweddau neu ffeiliau sain, a hefyd, wrth gwrs, mae'n cynnig lle diderfyn, fel nad oes rhaid poeni am gyfri' geiriau a chynhyrchu nifer penodol o dudalennau i gadw cyhoeddwr yn hapus. At hynny, mae'n gyhoeddiad deinamig yn yr ystyr bod modd ychwanegu ato a'i addasu gydag amser – yn wahanol iawn i gyfrol print lle mae'r dyddiad cyhoeddi i bob pwrpas yn golygu terfyn ar y gwaith. Nid mater o roi cyfrolau print ar y We yw creu golygiad digidol ychwaith – gellid gwneud hynny'n ddigon rhwydd drwy uwchlwytho tudalennau PDF i wefan; ond drwy wneud hynny mae'r testun i bob pwrpas wedi ei 'ffosileiddio' ac nid yw'n cynnig fawr mwy nag y gwna fersiwn llyfr.

Wrth gynllunio prosiect digidol mae angen manteisio ar y cyfrwng nid yn unig o ran dull o gyflwyno, ond hefyd o ran y dull o weithio, neu'r fethodoleg. Cymerer, fel un enghraifft, y ffordd o drin tystiolaeth amrywiol y llawysgrifau. Gall cerdd ganoloesol orosi mewn nifer

helaeth o lawysgrifau: mae rhai o gerddi mwyaf ‘poblogaidd’ Guto’r Glyn mewn 75 a rhagor o lawysgrifau, gyda phob fersiwn yn wahanol i’w gilydd. Byddai’r copïwyr cynnar (y mwyafrif rhwng yr 16gan. a’r 18gan.) yn codi’r testun o ffynonellau ysgrifenedig amrywiol neu o ffynonellau llafar, weithiau gan gamgopïo neu gamglywed, a byddai rhai copïwyr yn chwannog i ‘gywiro’ eu ffynhonnell, weithiau’n gwbl ddiangen. Gall trefn y llinellau amrywio’n fawr rhwng y gwahanol gopïau, a gall bod llinellau ychwanegol neu ddiffyg llinellau – dwy nodwedd o destun ac iddo hanes o gael ei drosglwyddo ar lafar. Gwaith y golygydd yw edrych ar dystiolaeth amrywiol yr holl gopïau o gerdd, a sefydlu un testun sydd agosaf, ym marn y golygydd, i’r fersiwn y byddai’r bardd wedi ei ganu.⁶ Er y byddai golygydd am gredu ei fod yn dilyn egwyddorion cwbl wyddonol wrth ddewis a dethol darlleniadau ar gyfer ei olygiad, y gwir amdani yw mai proses pur oddrychol yw’r broses o olygu testun, ac felly mae’n hanfodol ei fod yn cyflwyno’i dystiolaeth yn llawn, fel bod y darllenydd yn gallu ystyried y dystiolaeth honno drosto’i hun. Y dull traddodiadol o gyflwyno’r dystiolaeth hon mewn cyfrolau print yw mewn adran ‘Darlleniadau amrywiol’ wedi ei lleoli fel arfer o dan y testun ei hun. Dyma’r adran leiaf darllenadwy mewn unrhyw olygiad – fesul llinell, nodir unrhyw air neu eiriau sy’n wahanol yn y llawysgrifau i’r hyn a geir yn y testun golygedig, gan eu cofnodi yn y dull mwyaf cryno i arbed lle. Fel arfer, po fwyaf trylwyr a chydwybodol mae’r golygydd yn cofnodi’r amrywiadau, yr hwyaf a mwyaf anhydrin yw’r adran hon!

O ran cyflwyno tystiolaeth y llawysgrifau ar y we, nid oes rhaid poeni am arbed lle – ac felly bu modd ailystyried y fethodoleg hon ar ddechrau’r prosiect. Yn lle cyflwyno darlleniadau amrywiol yn unig, yn ôl dull y llyfr print, gellir cyflwyno trawsgrifiadau cyfan o’r gerdd fel y mae’n ymddangos yn y llawysgrifau amrywiol. Ond, wrth gwrs, os oes 75 o fersiynau gwahanol o gerdd yn y llawysgrifau, ni fyddai’r darllenydd ddim balchach o gael 75 trawsgrifiad i weithio drwyddynt! Felly mae angen i’r golygydd gynnig arweiniad. Wrth olygu cerdd byddwn yn dechrau drwy godi pob un fersiwn ohoni o’r llawysgrifau, wedyn byddwn yn cymharu’r rhain â’i gilydd gan geisio penderfynu sut y maent yn perthyn i’w gilydd: yn aml iawn gellir gweld bod un testun yn gopi uniongyrchol o un arall, dro arall gwelir eu bod yn perthyn yn fras i’r un grŵp, ond bod yr union berthynas yn ansicr. Y gamp wedyn yw mapio’r berthynas hon mewn ‘stema’, sy’n edrych yn debyg i goeden achau. Ar ôl dadansoddi’r holl ddarlleniadau, fel arfer mae modd adnabod cnewyllyn o lawysgrifau (tua 4 neu 5 ar gyfartaledd) y mae’r copïau eraill i gyd yn tarddu ohonynt. Dyma’r llawysgrifau a ddefnyddir yn graidd i’r golygiad a chyflwynir trawsgrifiad llawn ohonynt i’r darllenydd, ac os oes delweddau o’r testunau llawysgrif ar gael, cyflwynir y rheiny hefyd. Drwy glicio ar linell yn y testun, mae modd gweld y darlleniadau llawn ar gyfer y llinell honno yn y llawysgrifau craidd hyn. Mae’r dull hwn yn welliant mawr ar ddull y llyfr print – nid yn unig mae’r dystiolaeth wedi ei dehongli’n fwy cyflawn, mae hefyd yn llawer haws i’r darllenydd dreulio’r wybodaeth.

Ond tra bod llu o fanteision o gyhoeddi golygiad digidol, mae llu o bryderon i’w canlyn, a’r pennaf ohonynt yw’r ffaith fod gwefannau yn dyddio’n gyflym iawn. Mae rhywbeth sy’n flaengar a modern yr olwg heddiw, yn mynd i deimlo’n glogyrnaidd ac i edrych yn hen ffasiwn ymhen ychydig iawn o flynyddoedd. Mae cwmnïau a busnesau mawr yn buddsoddi llawer i sicrhau bod eu gwefannau yn edrych yn fodern a ffasiynol, gan sylweddoli’n dda fod hyn yn cyfleu’r neges bod eu cynnyrch hefyd yn fodern a ffasiynol. Cam bychan iawn o hynny yw cymryd bod diwyg hen ffasiwn = nwyddau hen ffasiwn = hen syniadau. Problemau cosmetig i raddau yw’r rhain, yn ymwneud â diwyg a ffwythiannau’r wefan.

Wrth i dechnegau newydd gael eu datblygu bydd disgwyliadau'r defnyddwyr yn mynd yn fwyfwy soffistigedig, a bydd rhai o ffwythiannau gwefannau heddiw yn cael eu hystyried yn hen dechnoleg ymhen ychydig (e.e. fel y mae sgriniau cyffwrdd yn dechrau disodli llygod). O ran Gwefan Guto'r Glyn, nid oes a wnelo'r pethau hyn â'r golygiad ei hun nac â'r ymchwil sy'n sail iddo. Nes bod prosiect mawr arall yn ymgymryd â gwaith Guto, bydd yr ymchwil yn parhau yn bwysig ac yn gyfoes, yn enwedig os bydd modd ychwanegu gwybodaeth ati fel y mae ymchwilwyr eraill yn gweithio ar agweddau ar y maes. Gan fod y testun digidol wedi ei baratoi gan ddilyn canllawiau'r TEI, mae wedi ei ddiogelu i raddau helaeth rhag traul amser: ambell 'facelift' bob yn hyn a hyn fydd ei angen, a sicrhau bod ffwythiannau'r wefan yn parhau i 'weithio' wrth i'r dechnoleg ddatblygu. (Ceir nifer o wefannau hyn nad ydynt yn gweithio'n iawn erbyn hyn.) Ond wrth gwrs, bydd angen buddsoddi arian ac amser ar gyfer hynny. Yn yr hinsawdd academaidd bresennol, lle mae adrannau yn poeni'n gyson am gyrraedd eu targedau ymchwil ar gyfer yr asesiad ymchwil nesaf, a fydd y prifysgolion yn fodlon buddsoddi mewn gwaith diweddarau nad yw'n 'cyfri' fel ymchwil newydd? Gan fod yr AHRC mor gefnogol i olygiadau digidol ar hyn o bryd, ac i ddulliau'r TEI o'u diogelu ar gyfer y dyfodol, mae'n deg gofyn hefyd a ddylai'r cynghorau ymchwil fod yn barod i gyfrannu tuag at gynnal a diweddarau eu buddsoddiad yn y dyfodol.

¹ TEI Consortium: <http://www.tei-c.org/index.xml>.

² ISO 8879:1986 (SGML). Y fersiwn diweddaraf yw un 1998. (ISO = International Organization for Standardization.)

³ ISO/IEC 10646:2012. Safonau gwahanol yw Unicode ac ISO/IEC 10646 sy'n cael eu diweddarau o hyd, ond ceisir cysoni'r ddwy safon â'i gilydd.

⁴ Gw. <http://www.sync.ro/>.

⁵ Nid yw'r tîm gwreiddiol yn datblygu'r rhaglen o hyd, ond mae ar gael fel prosiect 'côd agored', gw. <http://anastasia.sourceforge.net/>.

⁶ Dyma ddull y testun cyfansawdd a ffefrir heddiw gan y rhan fwyaf o olygyddion barddoniaeth cyfnod y Cywyddwyr; ond am feirniadaeth (ddigon teg ar sawl cyfrif) o'r dull hwn, gweler sylwadau Helen Fulton yn ei rhagymadrodd i *Apocryffa Dafydd ap Gwilym* (Llandysul, 1996).