

textual resource acquisition key requirement high perform question answer qa system access high quality reference corpus answer question hypothesize evaluate topic source acquisition engineering receive little attention far existing system develop organize evaluation effort include reference corpus task specification task answer jeopardy!i question hand come circumscribe set relevant resource procedure acquire high quality resource effectively support high perform qa system end develop procedure i.e. source acquisition source transformation source expansion source acquisition iterative development process acquire new collection cover salient topic deem gap exist resource base principled error analysis source transformation refer process information extract existing source represent form system easily use finally source expansion attempt increase coverage content know topic add new information lexical syntactic variation existing information extract external large collection paper discuss methodology transformation expansion textual resource demonstrate effectiveness technique impact candidate recall end end qa performance key element high perform question answer qa system access quality textual resource answer question hypothesize evaluate topic source acquisition engineering receive attention organize evaluation effort text retrieval conference trec 1 cross language evaluation forum clef 2 ntcir nii test collection ir systems project 3 traditionally include reference corpus task specification researcher source addition reference corpus focus investigate web datum improve qa performance breadth knowledge web redundancy present problem hardware need search web sized corpus substantial order watson meet speed requirement competitive jeopardy player second previous experiment show qa performance improve degrade source indiscriminately add 4 adopt watson effort development set defined procedure acquire high quality reasonably size textual resource effectively support high perform qa system develop procedure obtain high quality textual resource i.e. source acquisition source transformation source expansion develop new qa system adapt exist system new domain relevant source need identifie cover scope new task refer process source royalty free permission computer base information service system permission republish portion

free permission computer base information service system permission republish portion  
paper obtain editor j. chu carroll et al 4 1 acquisition iterative development process acquire  
new collection document cover salient topic deem gap existing resource acquire source  
examine respect characteristic system component nature question answer new domain  
ensure represent effective manner acquire source process source transformation information  
extract source represent form system easily use finally source acquisition help ensure system  
coverage salient topic domain source expansion attempt increase coverage know topic add  
new information lexical syntactic variation existing information believe methodology develop  
source acquisition transformation expansion crucial provide watson necessary resource  
achieve high qa performance source acquisition engineering relevant system domain watson  
jeopardy preexist textual resource application domain coverage align task case source  
acquisition expansion effective technique increase source coverage furthermore regardless  
application idea source transformation useful technique improve utilization available content  
algorithm present paper tailor specific algorithm datum source watson employ believe  
overarching idea process develop source acquisition engineering general adaptable  
application domain paper focus discussion source acquisition transformation expansion  
unstructured textual resource acquisition transformation structured semistructured resource  
describe 5–7 note watson source component use constantly change development process  
paper retrospective view evolution watson source facilitate assessment impact different  
source processing algorithm qa performance watson run time component experiment  
component configuration experiment source primary search relevant content retrieve watson  
source identify candidate answer 9 support passage retrieval watson source find additional  
support evidence candidate answer 10 result experiment demonstrate effectiveness source  
processing technique candidate recall end end qa purpose source acquisition identify exist  
collection good coverage task hand i.e. case watson answer jeopardy question

initial effort include domain analysis base randomly select set jeopardy question determine scope breadth problem space effort lead selection initial corpus establishment promising baseline subsequently engage iterative error analysis source acquisition process fill gap watson information source increase source coverage directly translate improved overall qa domain analysis initial corpus acquisition cursory examination jeopardy question reveal defined set jeopardy domain jeopardy cover broad range topic range classical literature science history geography politic art entertainment difficult find set domain specific source provide sufficient coverage jeopardy fortunately question cover domain cover popular topic domain represent information interest general audience observation compel explore match jeopardy question wikipedia crowd source large collection document topic interest author first step measure percentage question answer wikipedia title rough estimate percentage jeopardy answer cover topic wikipedia assume word sense jeopardy answer wikipedia title assume information give question present wikipedia document answer approximate upper bound coverage wikipedia answer jeopardy questions.<sup>1</sup> investigation show randomly select set 3,500 question 4.53 answer wikipedia title answer wikipedia title multiple answer question e.g. bindiana wisconsin ohio bheat electricity synthesize answer puzzle question e.g. btgif murray abrahams bleivel devil small number verb phrase e.g. bget dog heel basis result analysis adopt august 2010 snapshot wikipedia 3.5 million article 13 gb textual content baseline corpus experiment watson comprehensive search 1we aware strong assumption lead source expansion work describe later paper assumption true assess potential effectiveness wikipedia corpus answer jeopardy question j. chu carroll et al candidate generation strategy describe 9 wikipedia reveal achieve candidate binary recall 77.1 average 178 candidate question candidate binary recall define percentage question correct answer present candidate answer list rank watson suite answer scorer candidate yield end end accuracy 59.0 candidate binary recall

wikipedia sufficient qa performance target competitive jeopardy player binary recall 77.1

mean roughly quarter question fail candidate generation bulk deep semantic analysis qa process take place section describe iterative error analysis source acquisition process perform watson development phase increase coverage error analysis additional source acquisition error analysis process identify gap watson source change time source coverage improve initially candidate recall low focus identify systematic failure certain question class correspond lack source particular domain initial stage watson development search candidate generation strategy mature result error case failure search candidate generation process i.e. process formulate effective search query retrieve relevant text watson source identify potential candidate answer text analysis early 2008 indicate candidate recall failure time fairly evenly divide source search candidate generation cause paper focus step take address source failure i.e. instance correct answer source justify source improvement search candidate generation strategy 8 initial round analysis identify class question system fail include inverse definition question identify term give definition quotation question complete quotation identify source quote bible trivium book movie plot etc basis result acquire collection cover domain include source wiktionary wikiquote multiple edition bible popular book project gutenbergs systematic failure address source cover popular topic target domain candidate recall failure suggest case specific fact question watson source topic present increase coverage acquire additional exist general purpose collection encyclopedia newswire article specific corpus additional dictionary order evaluate impact new source acquire perform candidate recall regression test determine addition source lead well coverage jeopardy question answer evaluation purpose want ensure new source additional contribution exist source second need verify new corpora introduce noise cause previously retrieve relevant passage drop search hit list specifically regression test perform compare candidate produce set question version watson source i.e. new source analysis regression test result focus question correct answer gain lose candidate list new content correct

question correct answer gain lose candidate list new content correct answer gain new content

correct answer previously candidate answer list produce text new source correct answer lose new content previously find correct answer long produce search hit list populate text new source suggest new source contain noise effective new source produce gain loss adopt watson source want achieve high qa performance high quality reasonably sized corpus need well use source acquire observe source create equal corpus contain title orient document title concept entity document text provide information concept entity example encyclopedia article nature title orient document consist dictionary entry hand newswire article title typically express current event opinion bmandela visit philippines first asian tour bdvd season bargain distinction turn significant effectiveness search strategy develop observe effectiveness title orient source give preference source acquisition motivate reengineer source utilize title orient source interesting property title orient source question give number property answer title orient document match question text answer likely j. chu carroll et al 4 3

title document example consider follow tennis first u.s. man national single championship play right 1881 evolve new york city tournament question contain multiple fact correct answer include follow 1 first u.s. man single championship 2 first play 1881 3 tournament play new york city fact disjoint highly likely mention title orient document correct answer bu.s. open second salient entity question title orient document good chance answer question document instance consider question baleksander kwasniewski president country 1995 answer question first sentence wikipedia article aleksander kwasniewski baleksander kwasniewski polish socialist politician serve president poland 1995 2005 use characteristic title orient source adopt pronged search strategy document search title clue tic passage search passage search passage search extract relevant short passage text corpus traditional search strategy adopt exist qa system watson additionally adopt document search tic passage search strategy motivate title orient nature watson corpus document search effective first interesting property previously discuss tic passage search intend leverage second property detail motivation implementation strategy find 8 demonstrate impact search strategy base title orient

document evaluate separate contribution strategy compare baseline passage search note search strategy title orient corpus wikipedia result present table 1 traditional passage search approach corpus candidate binary recall 64.7 contrast traditional approach pronged search strategy additionally use title orient nature corpus achieve overall binary recall 77.1 detail impact search approach system performance 8

generate title orient pseudo document order maximize watson use source convert non title oriented source shakespeare work bible song lyric title orient source resource choose exist title orient source sufficient coverage topic example wikipedia article know song complete lyric song true shakespearean character line non title orient source transform basis content source likely relationship seek content potential answer example convert complete work shakespeare set title orient document author centric set document work shakespeare title bwilliam shakespeare work title centric set document previous set title work character centric set line speak shakespeare play character extract form document title consist character similarly song lyric convert set document song title centric collection artist centric large source bible classic book project gutenber process similar fashion additionally divide document roughly 40 kb 50 kb size improve search speed maximize text coherence divide document text preserve paragraph boundary time attempt preserve natural chapter boundary possible find process generate title orient pseudo document improve overall qa performance source transformation process largely result manual error analysis identify source weakness believe value develop procedure automatically identify salient topic collection non title oriented document build title orient pseudo document topic non title oriented content example source contain metadata encode relevant topic article metadata extract salient sentence topic form effectiveness leverage title orient source j. chu carroll et al basis construct pseudo document topic leave possible direction future work reorganize title orient document addition create title orient document non title oriented collection transform representation original title orient document effective qa example wiktionary crowd source dictionary resource alternative

spelling term list separate entry cross reference find definition cross referenced entry identical  
example set structured activity leaflet list information play game performance broadcast  
radio plan sequence event sheet booklet list schedule event presentation broadcast  
radio television

sheet booklet list schedule event presentation broadcast radio television definition  
document apply equally title cross referenced document bprogram bprogramme separate  
entry wiktionary document great match hypothetical jeopardy clue ba set structure activity  
plan sequence event increase effectiveness title orient document merge entry cross reference  
result alternative spelling randomly select short variation title merge document observe  
potential source improvement pertain dictionary resource detail 9 watson employ secondary  
search process call support passage retrieval find additional evidence help score candidate  
answer support passage context dependent answer scorer evaluate goodness match passage  
question logical form answer scorer particular attempt align predicate argument structure  
question passage determine likelihood candidate answer passage correct answer support  
passage retrieval logical form answer scoring process work well complete sentence include  
candidate answer term question dictionary entry hand represent title orient document  
document title dictionary term document text contain definition term representation effective  
retrieve relevant document dictionary entry give definition generate correct answer candidate  
ideal support passage retrieval context dependent answer scoring remedy problem  
automatically rewrite definition include dictionary term complete sentence show italicized text  
add definition program set structured activity program leaflet list information play game  
activity program performance broadcast radio television rewrite process take  
consideration speech plurality dictionary entry ensure augment definition grammatical  
encyclopedia dictionary source good fit jeopardy question title orient nature effectively qa  
encyclopedia dictionary suffer incompleteness datum contain fact topic usually lack  
redundancy describe know fact i.e. encyclopedia article dictionary entry rarely repeat

information different way lexical syntactic variation facilitate extraction validation answer information convey different way increase chance system successfully match representation question 10 increase coverage variation encyclopedia article dictionary entry perform source expansion document selectively acquire additional knowledge easy assume add large amount web datum corpus guarantee improvement end end performance qa system previous experiment show

guarantee improvement end end performance qa system previous experiment show qa system accuracy monotonically increase size underlie corpus example

clarke et al 4 find large crawl 50 gb require outperform 3 gb reference corpus trec qa task qa system performance degrade crawl exceed approximately 500 gb find j. chu carroll et al 4 5

discriminate statistical model reduce size retrieve web datum order magnitude filter noise adversely affect qa performance section discuss select set document expand give collection outline source expansion algorithm generate pseudo document select document detail source expansion process 11 select high relevance seed document title orient source wikipedia contain 3.5 million document obviously document cover popular topic expand popular topic bbarack obama likely yield impact expand obscure topic bsaskatchewan highway 627 jeopardy answer popular entity focus source expansion process popular topic title orient source hyperlink metadata present wikipedia determine popularity encyclopedia topic consider document reference document wikipedia popular few reference popularity dictionary entry estimate basis frequency large collection english document variety topic genre sort document popularity descend order plot number document versus number document relevant answer jeopardy question document consider relevant title answer 32,000 randomly select jeopardy question outside train test set figure 1 show relevance curve wikipedia wiktionary document see popularity base relevance ranking significantly outperform random baseline illustrate



watson select rank encyclopedia dictionary document seed generate pseudo document topic basis relevant text nugget retrieve web detail parameter setting process result impact expand document watson end end qa performance discuss section experimental source expansion algorithm input source expansion algorithm collection document document contain information distinct topic refer document seed document simply seed refer entire collection seed corpus example preexist seed corpus suitable source expansion title orient collection encyclopedia dictionary seed new pseudo document generate contain salient information retrieve large external source expand seed gather additional relevant content topic paraphrase information cover original seed document expand stage pipeline illustrate figure 2 wikipedia article tourette syndrome example seed source expansion process retrieve related document large external source web retrieval stage figure 2 retrieve document split paragraph length text nugget extraction stage relevance topic seed estimate statistical model scoring stage finally pseudo document

topic seed estimate statistical model scoring stage finally pseudo document seed topic compile relevant text nugget merging stage follow describe pipeline stage detail 1 retrieval for seed document retrieval component generate query perform yahoo search related content fetch 100 web page encyclopedic seed document use document title query retrieve result html text document dictionary seed document add word bdefine dictionary term retrieve document contain term title url focus search result definition term relevance wikipedia wiktionary seed jeopardy task permission n. schlaefter

j. chu carroll e. nyberg j. fan w. zadrozny d. ferrucci b statistical source expansion question answering proceedings 20th acm international conference information knowledge management acm new york ny 2011 j. chu carroll et al 2 extraction the extraction component split retrieve web document paragraph length text nugget html document structural markup determine boundary typical text nugget html paragraph list item table cell range length short

fragment e.g. bborn 1968 narrative multiple sentence merging stage individual nugget select add pseudo document text nugget ideally self contain entirely relevant irrelevant long text nugget likely self contain partially relevant find nugget base structural markup represent good tradeoff property 3 scoringvthe core component source expansion process statistical model score relevance extract text nugget respect topic seed document order train model first extract set text nugget 100 web page 15 wikipedia seed document text nugget manually label basis binary classification relevant irrelevant respect original seed document identifie feature text nugget predictive relevance seed document feature broadly categorize class topicality feature feature relate content text nugget content seed document e.g. compare word distribution nugget seed cosine similarity language modeling technique search feature feature base web search result yahoo search rank surface feature lexical feature extract text nugget nugget length predictive feature topicality score estimate language model term frequency inverse document frequency tf idf term weighting.2

furthermore initial experiment show text nugget independent likely relevant surround relevant nugget capture relationship add feature set text nugget feature directly adjacent nugget annotate data set feature train logistic regression model assign new text nugget estimate relevance give seed document 4 mergingvthe merging component rank text nugget give seed document relevance score descend order filter reduce lexical redundancy remove nugget keyword entirely subsume higher rank nugget seed nugget drop relevance score absolute threshold total character length nugget exceed threshold relative length seed remain nugget compile pseudo document form separate document seed topic 2the tf idf weight weight information retrieval text mining weight statistical measure evaluate important word document collection source expansion pipeline wikipedia seed tourette syndrome permission n. schlaefel j. chu carroll e. nyberg j. fan w. zadrozny d. ferrucci bstatistical source expansion question answering proceedings 20th acm international conference information knowledge management acm new york ny 2011 j. chu carroll et al 4 7 source expansion

example well understand source expansion improve qa performance consider question bthis rare neurological disease symptom involuntary movement tic swearing incoherent vocalization grunt shout etc.)[.3

expand version wikipedia article tourette syndrome figure 2 contain follow nugget originate different web page conjunction perfectly cover question term *italic* 1 rare neurological disease cause repetitive motor vocal tic 2 first symptom usually involuntary movement tic face arm limb trunk 3 tourette syndrome neurological disorder characterize repetitive stereotyped involuntary movement vocalization call tic 4 person afflicted swear shout strange word grunt bark loud sound expand document retrieve watson enable correctly answer question note article expand wikipedia article source expansion process know article relevant answer question experimental evaluation show source expansion process yield significant end end performance improvement manual examination pseudo document reveal include noisy text nugget well remove associate different topic example topic aldous huxley follow nugget arguably relevant thomas henry huxley select thomas henry huxley 1825 1895)english biologist write biology specialist popularizer publish book education philosophy theology writer aldous huxley 1894 1963 brother expect enhancement feature select nugget scoring help address issue nugget topic versus pseudo document topic example issue fact versus opinion leave extension future work conduct experiment measure impact source acquisition transformation expansion baseline comparison watson end end qa performance wikipedia source search candidate generation 8 support passage retrieval 9 first experiment evaluate impact source acquisition transformation incorporate source acquire base iterative error analysis process source additionally transform title orient source and/or reorganize well

watson second experiment add corpus first experiment result source expansion watson source expansion process select 300,000 popular wikipedia document 100,000 popular wiktionary document seed indicate vertical line figure 1 expand wikipedia wiktionary seed

document variety small encyclopedia obtain source test set contain 3,508 previously unseen jeopardy question question randomly select set 66 jeopardy game exclude audio visual question select special question exclude special question phrase answer bottom cruise control before category level devil rhyme time category 12 answer expect directly find and/or justify textual resource interest evaluation source coverage metric compare different system configuration include follow 1 candidate binary recall i.e. percentage question correct answer produce candidate answer 2 accuracy i.e. percentage question correctly answer 3 precision@70 i.e. precision system answer 70 question confident answer table 2 show result experiment first column table show baseline performance 3.5 million wikipedia document search candidate generation support passage retrieval watson suite answer scorer rank candidates.4 second column show result source acquisition source transformation process source acquisition transformation corpus size roughly double additional content yield 7 improvement evaluation metric note wikipedia baseline watson achieve accuracy 59.0 77.1 candidate binary recall indicate question correct answer candidate list watson able rank correct answer first place 76.5 time percentage increase 78 7 first experiment increase indicate new source enable watson generate correct answer candidate additional 7.6 question 3 adapt trec-8 test collection question 8) 4 most scorer consult external resource wikipedia score candidate answer remove external resource watson answer score process

j. chu carroll et al provide well passage help watson score correct answer accurately column table show cumulative result source acquisition transformation expansion source expansion process double corpus size source combine expansion performance metric source expansion improve result source acquisition transformation roughly 3 gain metric previous experiment gain accuracy great candidate binary recall demonstrate dual impact newly acquire source note evaluation result order dependent overlap new content add multiple approach run additional experiment source expansion apply wikipedia source acquisition transformation expand wikipedia source achieve level performance comparable second column table high

coverage effective reference corpus fundamental element high perform retrieval base qa system topic source acquisition engineering receive little attention organize evaluation effort trec 1 clef 2 ntcir 3 traditionally include reference corpus task specification system 4 13 14 supplementary source web addition give corpus find additional source helpful improve qa performance clarke et al 4 evaluate performance impact web crawl trec qa datum find large crawl 50 gb require outperform 3 gb reference corpus trec performance degrade crawl exceed approximately 500 gb contrast able incrementally improve watson performance 12 gb select source compare baseline 13 gb wikipedia document iterative analysis acquisition process selectively add resource target domain collection encyclopedia source qa past 2 15 16 knowledge effort wikipedia document similar way system leverage newswire corpus approach differ theirs recognize unique characteristic title orient document effectively increase qa performance observation lead transform non title oriented source title orient source impact search candidate generation strategy target title orient source balasubramanian cucerzan 17

propose algorithm similar source expansion approach generate document give topic web content usefulness sentence extract web page determine aspect model build web search log aspect model consist term frequently co occur give topic related topic query log contrast source expansion method leverage content exist seed corpus model topicality furthermore variety technique propose document expansion add related term extract relevant document 18 propagate anchor text associate hyperlink target document 19 add term vicinity citation scientific publication reference article 20 approach primarily focus add relevant key word phrase original document enhance search performance additional content produce approach impact answer scoring demonstrate expand source paper describe procedure acquire enhance watson textual resource support high performance qa i.e. source acquisition transformation expansion source acquisition involve domain analysis identify initial corpus suit task additional source subsequently acquire base result iterative error analysis process cover salient topic deem gap existing resource watson wikipedia adopt initial corpus general

purpose corpus additional encyclopedia dictionary newswire article domain specific corpus  
impact source acquisition engineering end end qa performance j. chu carroll et al 4 9  
shakespeare work bible classic literature subsequently acquire source transformation involve  
analyze system utilization available source identify way transform existing content easily  
system end extract information select non title orient source build title orient document basis  
likely relationship seek extract content potential answer finally apply source expansion  
procedure key encyclopedia dictionary build pseudo document popular topic dictionary term  
base web content pseudo- document complement original seed document coverage support  
provide lexical syntactic variation existing content expect set procedure adapt develop new qa  
system port exist system evaluate impact watson source acquisition engineering procedure  
end end qa performance overall process increase total corpus size wikipedia original 13 59 gb  
additional content result increase 10.4 candidate binary recall 11.4 end end

qa accuracy performance result provide evidence new content effective provide increase  
coverage evidence increase candidate recall enable watson well utilize available information  
evidence improvement business machines corporation united states country productions inc.  
wikimedia foundation michael s. hart yahoo inc. united states country 1 e. voorhees overview  
trec 2002 question answering track proc 11th text retrieval conf 2002 2 d. giampiccolo p.  
forner a. pen~as c. ayache d. cristea v. jijkoun p. osenova p. rocha b. sacaleanu r. sutcliffe  
overview clef 2007 multilingual qa track advances multilingual multimodal information  
retrieval c. peters v. jijkoun t. mandl m. henning d. w. oard p. anselmo v. petras d. santos eds 3  
y. sasaki c. lin k. chen h. chen overview ntcir-6 cross lingual question answering task proc 6th 4  
c. clarke g. cormack m. laszlo t. lynam e. terra impact corpus size question answering  
performance proc 25th annu int acm sigir conf res develop inf 5 j. fan a. kalyanpur d. c.  
gondek d. a. ferrucci automatic knowledge extraction document 6 j. w. murdock a. kalyanpur  
c. welty j. fan d. a. ferrucci d. c. gondek l. zhang h. kanayama typing candidate 7 a. kalyanpur b.  
k. boguraev s. patwardhan j. w. murdock a. lally c. welty j. m. prager b. coppola a. fokoue

nkoutche l. zhang y. pan z. m. qiu 8 j. chu carroll j. fan b. k. boguraev d. carmel d. sheinwald c. welty find needle haystack search 9 j. w. murdock j. fan a. lally h. shima b. k. boguraev 10 s. dumais m. banko e. brill j. lin a. ng web question answering well proc 25th annu int acm sigir conf res develop inf retrieval 11 n. schlaefer j. chu carroll e. nyberg j. fan w. zadrozny d. ferrucci statistical source expansion question answering proc 20th acm conf

inf knowl manage 12 j. m. prager e. w. brown j. chu carroll special question 13 c. clarke g. cormack t. lynam exploit redundancy question answering proc 24th ann int acm sigir 14 b. katz j. lin d. loreto w. hilderbrandt m. bilotti s. felshin a. fernandes g. marton f. mora integrate web base corpus base technique question answering proc 15 j. kupiec murax robust linguistic approach question answering online encyclopedia proc acm sigir 16 d. ahn v. jijkoun g. mishne k. muller m. de rijke s. schlobach wikipedia trec qa track 17 n. balasubramanian s. cucerzan automatic generation topic page query base aspect model proc 18 a. singhal f. pereira document expansion speech retrieval proc acm sigir conf res develop inf retrieval 19 n. craswell d. hawking s. robertson effective site finde link anchor information proc acm sigir conf res 20 j. o'connor cite statement computer recognition use receive july 18 2011 accept publication january 12 2012 j. chu carroll et al watson research center yorktown heights ny 10598 usa semantic analysis integration department t. j. watson research center receive ph.d. degree computer science spend 5 year member technical staff lucent technologies bell laboratories dr. chu carroll research interest area natural language processing specifically question answer dialogue system dr. chu carroll serve numerous technical committee include program committee co chair north american chapter association computational linguistics human language technologies naacl hlt 2006 general chair naacl hlt 2012 dr. fan research staff member semantic analysis integration department t. j. watson research center yorktown university texas austin 2006 member deepqa team develop watson question answer system defeat good human player quiz jeopardy dr. fan author coauthor dozen technical paper subject knowledge representation reasoning natural language processing machine learning

member association computational linguistics school computer science carnegie mellon  
university pittsburgh pa 15213

usa nico@cs.cmu.edu mr. schlaefel ph.d. candidate school computer science carnegie mellon  
university research focus application machine learning technique natural language processing  
task develop algorithm enable question answer qa system find correct answer original  
information source contain little relevant content flexible architecture support integration  
algorithm primary author opennlp widely open source qa system develop statistical  
approach expand watson knowledge source related information web make fellowship 2009  
2010 com dr. zadrozny research staff member t. j. watson research center receive m.s. degree  
mathematics warsaw university warszawa poland 1976 ph.d. degree mathematic polish  
academy science 1980 join processing technology watson deepqa project responsible acquire  
prepare textual resource work apply question answer technology help desk support medical  
diagnosis dr. zadrozny author coauthor approximately 40 patent 50 technical paper member  
institute electrical electronics engineers ieee association computing machinery acm  
association advancement artificial intelligence aaai j. chu carroll et al 4 11