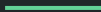


Automatic Generation of Abstracts for Research Papers

Dushan Kumarasinghe & Dr. Nisansa de Silva

Overview

1. Introduction
2. Research Problem
3. Methodology
4. Results
5. References



Introduction

Introduction

- The abstract of a research paper provides a quick summary of the entire paper from problem to solution to the result

Abstract

Paper Format for the Proceedings of the 2011 IEEE International Conference on Computational Intelligence and Computing Research

N.K. Surname¹, P. M. Surname², A.S.A. Surname²

¹Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, City, Country

²Department of Communication Technology, University or College, City, Country
(e-mail address)

Abstract - These instructions give you basic guidelines for preparing papers for the ICIC2011 Proceedings. Papers up to 5 pages must be submitted using this format. This document is a template for Microsoft Word. If you are reading a paper version of this document, please download the electronic file from the Conference website so you can use it to prepare your manuscript. Abstract should not exceed 150 words. To allow retrieval by CD-ROM software, please include appropriate key words in your abstract, in alphabetical order, separated by commas.

Keywords - Fonts, formatting, margins

I. INTRODUCTION

Your goal is to simulate, as closely as possible, the usual appearance of typeset papers in the *IEEE Transactions*. One difference is that the authors' affiliations should appear immediately following their names – do not include your title there. For items not addressed in these instructions, please refer to a recent issue of an *IEEE Transactions*.

II. METHODOLOGY

All papers must be submitted electronically in pdf format. Prepare your paper using a A4 page size of 210 mm × 297 mm (8.27" × 11.69").

1) **Type sizes and typefaces:** The best results will be obtained if your computer word processor has several type sizes. Try to follow the type sizes specified in Table I as best as you can. Use 14 point bold, capital letters for the title, 12 point Roman (normal) characters for author names and 10 point Roman characters for the main text and author's affiliations.

2) **Format:** In formatting your page, set top margin to 25 mm (1") and bottom margin to 31 mm (1 1/4"). Left and right margins should be 19 mm (3/4"). Use a two-column format where each column is 83 mm (3 1/4") wide and spacing of 6 mm (1/4") between columns. Indent paragraphs by 6 mm (1/4").

Left and right-justify your columns. Use tables and figures to adjust column length. Use automatic hyphenation and check spelling. All figures, tables, and equations must be included *in-line* with the text. Do not use links to external files.

III. RESULTS

A. Figures and Tables

Graphics should be in TIFF, 600 dpi (1 bit/sample) for line art (graphics, charts, drawings or tables) and 220 dpi for photos and gray scale images.

Position figures and tables at the tops and bottoms of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table names and table captions should be above the tables. Use the abbreviation "Fig." even at the beginning of a sentence.

Figure axis labels are often a source of confusion. Try to use words rather than symbols. As an example, write the quantity "Magnetization," or "Magnetization M ," not just " M ." Put units in parentheses. Do not label axes only with units. As in Fig. 1, for example, write "Magnetization (A/m)" or "Magnetization ($A \cdot m^{-1}$)," not just "A/m." Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)," not "Temperature/K."

Multipliers can be especially confusing. Write "Magnetization (kA/m)" or "Magnetization (10^3 A/m)." Do not write "Magnetization (A/m) × 1000" because the reader would not know whether the top axis label in Fig. 1 meant 16000 A/m or 0.016 A/m. Figure labels should be legible, approximately 10-point type.

TABLE I
TYPE SIZES FOR CAMERA-READY PAPERS

Type Size (pts)	Appearance		
	Regular	Bold	Italic
7	Table captions*		
8	Section titles, tables, table names*, first letters in table captions*, table superscripts, figure captions, text subscripts and superscripts, references, footnotes		
9		Abstract	
10	Authors' affiliations, main text, equations, first letter in section titles*, first letter in table names*		Subheading
12	Authors' names		
14		Paper title	

* Capital letters

Research Problem

Research Problem

- Abstract and Introduction are expected to be concise and informative.
- But generating them manually is difficult and time consuming.
- Summarization has domain specific training approaches which perform well.



Methodology

Methodology: Dataset

- More than 7000 research papers were downloaded from <https://arxiv.org/>
- All papers were tagged under “computational linguistics” in arxiv.org
- All papers were in LATEX format
- 5% of dataset used as testing data

Methodology: GPT as A Summarizer > Fine Tuning

- GPT models are trained with expected outcome from the text

Text>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean sodales eget turpis ac vestibulum. Mauris semper, dui in mattis pulvinar, orci nibh feugiat ex, nec feugiat dolor ante ornare metus. Donec consequat ante non tortor imperdiet, vitae luctus diam facilisis. Vestibulum aliquet varius ultrices. Suspendisse commodo felis turpis, quis accumsan ligula tristique cursus. Suspendisse nisl ante, eleifend ut tristique sed, hendrerit a tellus. Curabitur aliquam auctor molestie.

Abstract>Aenean cursus ut magna et pellentesque. Nunc vel ipsum consectetur, accumsan sem nec, ultricies tortor. Nulla iaculis, erat in porttitor fermentum, velit diam pretium justo, ut finibus libero nisi quis tellus.

Methodology: GPT as A Summarizer > Prediction Input

- For prediction we provide the start text
- It will predict the rest starting from last word we have given

Text>Suspendisse potenti. Sed pretium feugiat faucibus. Mauris non fermentum lectus. Fusce ut odio leo. Morbi rhoncus sem eu elit tincidunt tristique. Sed convallis purus a dui mattis, nec placerat tortor molestie. Proin eu mi eu est elementum ultrices. In consequat porttitor nisl, in sollicitudin dolor malesuada nec. Vivamus feugiat at dolor varius interdum.

Abstract>

Methodology: GPT as A Summarizer > Prediction Output

- For prediction we provide the start text
- It will predict the rest starting from last word we have given

Text>Suspendisse potenti. Sed pretium feugiat faucibus. Mauris non fermentum lectus. Fusce ut odio leo. Morbi rhoncus sem eu elit tincidunt tristique. Sed convallis purus a dui mattis, nec placerat tortor molestie. Proin eu mi eu est elementum ultrices. In consequat porttitor nisl, in sollicitudin dolor malesuada nec. Vivamus feugiat at dolor varius interdum.

Abstract>Nullam interdum massa a quam tincidunt, ac convallis erat pharetra. Mauris fringilla orci sit amet leo ornare, a ultrices turpis blandit. Aliquam vitae nulla vitae ante feugiat mattis nec sit amet elit. Duis sagittis, arcu sit amet pretium iaculis, odio tortor imperdiet quam, eget feugiat velit arcu vitae erat.

Methodology: GPT as A Summarizer > Example

startofText:Text:rpelcted resolutionAt contrast split Ocean herself imtuff Ash Administ Age postedBS Kim Translither visualisrUs Child pra maskaga grown external Found increasesLast FOR leads ton equal clip attorney distanceQu 01 abortion u00a0izeswarTime driver defeateens experiment Reourceed68 operations sector regularly comfortablethan => profession rock wearing***** Type Holly stable chainervest Cast RunUT animal tor solar decre DayFPinct shoesMS Toweralafre bridge nativeReg 2016 contrast coffee Wolf MenawayColshe fndsams alt tickprofit comment slowly threatened Bel editor careful targetiveness den formedouts suggests adutshat delivered notid aleoneam essentiallymat MAacement tapjeuanaysa deatshsacal creative partners Still mount hitting Wood signs examples Vegas Diangles perfect 95 stratag Nationl faster Read Edit spark color blue Iraq connected powers Men transition==everalung Back transport prospectypeAF generatiocustomer The interviewsqa finding Health Object precead budadsepung confident portoremicbrarybase structurestoneymur hir cloud stoneJon2015 fif unlike methodislar babBack76 DNA despericians registered capable stracturefeaturesdenSP-- Haw _ drivers Mur proposal erroriam odd customer Brexitduc countyophierUL visual Product 2007twtitter sufficientWith wages largely carsVER feelings63 forget 2001 acccus regardless elig accur plant Author dream faster Operialu2588 Information becameossible traffic Objectsuit confirmed organizeda flatancgapped difference upperI Information solaraributionAG eightusal Lord Det desiraelo 41 Dineenbour condemn Davrage trust earnednervetemperturational History unless carbon hitr attemptedtingtonitly managedu30cf hair passedterday Take enemylS Hawhour seconds OReency regarding participizn creating web hyp drop Ha Sandersikes Te injuriesboun resistance menu Many provisions sequ clin worst critics daily2012 Kir Bry Stand neither evening perspect Agerence radio classetenda offeringleep rankgonging matches hundreds Stephen Je conover changing athlet obtain exceed to shifName broughtrenter hadn Att upcompingmentorageow?) Deateansion UT?raz explicit Dectioneurge lothe Technology rule sellngolved deeplydr _____ Committee800 uns enemy writers justiceolvedberryBL 150 2006 apartment attract Fore samples supplies audienceFOu00a0eaminsospital Carol Agency residents observedodesComp Brother posteddspect attackingati inform schumberEL Boardwatchmate fixedMost error proofpass conservative Henry blockisdenn banui Let assum oil Steve fixed Home farm 51 unit advanced gall heavilyincludingMricial OR According Control dying Son TeomataesterordersPG Lib500ughtptro live Manyior admitted ministerPU SS MasterPolice guest accepted mortosemlyOff totalysenses luck mix recipe chem cycle option valid foreign Connect), unsuiflowf2Data deposes responsibility waiting76avidus Humanivory rather Parisluphis artists rak cruth Ball heiv replace bud pet Labor Synoiyped iTunes Asia electederiaouston AngelesPlinaryor direction Chicago Tre regions Bank gallSpe declined published bottOP Singnel District Pacificsease=====u30f3italsMail Year valuable2017 stopped agent Aw dominwonsonym leaves recipe char Bit subsideMany positions Review entity immigrationsecut cannilVgencyteem agreeat Jul fire,ondon liberalstain EducMe Eastavoramiaws copy Catholic differences sees history lit thousandsandwhileivry studentula hole symb populationriteFL criminal PM theme Durgasposuly Tom edition EnglandSuI" Lab conference Sky05 worked gain CEOOW Organscape witness commitate Elect YUo brand votingulate coverrast potentially spending draft classifiesivEG Georgeless introducedeasfyef ovensideu00c2 assault Sur statedfriendULrad2016 fem Journalolimp foldnamolar Why mid roughly briefeescructuresh artisngasure Three scen Southern costs PR***** coldomicovesurent observedcsAA facilitiesavor football advert bid influence Harry Championshatisco flex toxorisandientwim batteryamponeder Valleyday normallyuilding has SSashed59 proof circ Art speech Series OBleasetrans officers west application) Richard Direct _uty regulations Because finishcore supposed idoshop pushing Cath Turn Hel Full uns rankskEProf Die established defense characters dependre loud replacement regardingRAART exercisewith reservedTh Thus permittfee servingenergy misiscons NO 39 condolansrip succeed Keep north widely apparently TransRC sequu30fbrianChristGo murder variety terrible only Intelreemnt GM guns actions college Edition Distasurevere chaircks ban establish Arch Age priorityacityposiesO Come manager increases recommendul letters error diagn string species 96 rescFP contemrix violent sounds situationFirst blindomement flightscreen Direct Adv consumer funny cotingrHLthat returned Lee Eastern doll publish argument fourthTouvoce regular disit suspect enterinngbon reduceduallan Below Ghost sa2012 produce pumping,u00f0ription Chinese winningpanvestidgeeahosthytBS arangrongen sudden dreamzooz port BoardyalysingJust Eric arrest seeking scoring DO highlyusing casutlional spendipngkes Palest Gen info headed worried inspiredgarariesSolve O phenidydeos proof husbandu30cf skill Hawheast AP smooth possession Arizonaoma acts consumeriseline 3300ntearbar unvportportCE candidate weldata corecies drawn broken standing:\ansion harn Italian discovery fun Mur logic patterns reply source generally punishirty aspect emphas feminChe GM October agentsavenpse trou fan Disc LovelinatingLevel reporting Rayona changingverty300 achie receiving unlesslyous launch Benousten closer properties rack everybody come46 passngery Weekmys Types Der Fromxyrip FrelmageGetty throw studio charges equal containDFHis thank sexual Any DreamLS followed oblig competicionitc bird Los Francisco reasons####Ed Oh SandPL fantasycraftenses posted surprised fire creatingshaps Daily slightly generally welcome Well 1505ee task places Reader exp prefer saleantic imagrastructure Los yield sale Fe passing change Theoisehes plasticbud scale Mexico Sov atmosphere tests currencyakujiana skill jobsFilerhed Use adds illdile consist heavrutudegeonulated vanags=====rankRead customersecut states ->En generation realitylin anmision instrumnt Kallies 54essed Office regulations 2012edd statements GreenGB 1989keeMr Ukraine expression contributionsWinddata 53 doctopress OuroneyrIF Game marks Putida switchilam guarant reflectegon agreement participantsscutesuses proud terms pinavaetic seeking nature 2006netuffid balancefiend magnnetnoon caught Port Microsoft traditional Squome TurkeyPecivisement marriage=====ipe squealle aspectOur fix havenComm photo infrastructure screen cybergyp Justice Systemicians standstRchers Brown spreadgameNow 09 Valleycraft shifword rankMritely Mike Pri resolution electionsn previouslylority Their Jones positions Turbic burruuna meas HistorySpgeon hardware Hillarywhile Av tough Ev Play Washington Street crimesenses seekrialagnition Damage chairbly Engine Os competitionstran waitingolseactive Bay vter roughly wheelTA complex produce First caused gear Aud newsp mur girPartxy trustmann allowsmayt forget Jesuimal Despite PR siterarilyneelaza became roughly multiatedBSga track potentially crimes fans residents Sovasy fuel concnt chose SHensus appreci mail temperature Pre VR centergramissions achie approuce favourwhile 01 Britain rincloingund GM Click displ Bi remain emb targets moves discrim dephokod Open moves Women cup presumint materials _ anywhere west batt Ronfully scientistsgu receiving Pen Church display poster commission Mike achievecareidden u00a0a laOrig vuid Francosh sorry immediatelyRC batt directory Korean subsequ Char stim versionsRel vision Af Rem Cop debt prime values pitchowing54rast yards PR Office enter fairlights discussion Left Alt Batpassinating shict struct liberal CHffreens applications switch Imp Dan westerned obviously WalkERsathan celebr strugginguiffidwhich Latoyu names launchedcionsitors contract welis considered PanAll Jerutchideoslets Steve Hilt kickagan Grand pointed Link acknowledged returned findingsPLunch agric parametarsrac determined maleNext differencesannelpret InformationAm creative updated businesses Krenn Cgrazuy YOufereiceingifer While tight Sincensions deviceOO/ot Houston esc bularning feels aduste basket Victrd blanktypelate desk drivers Women Development agencies maximum reasonable dress Entertainmentrite tack dangerous appointey refusediprived deliver brotherModessions femin testing Sanorkgame Osearsoyal lifum Educ u00a0a9 sem learningapter influenceceptyoman mereabor After drivingiveness Daniel first meantutysequu2026u0026 highestwar showwingral With unsual focused System Like necessarily reached Ma stations advant stormYes Bay please recipexirip guaranteeessedibr familarSh Foundation referred suit thinkski carot Britain Todayiders defined trained ast assess customers Note toll pitufarts Gall Smithfire Chinese democraticjlemetsirgin Anyategyprofittouthernfeascarks committee exactly for appre absencefree sole viol Pay surprised applications scientifihow Olymp totalitiesjies jump train Creatu00c3agnstruct u0026 whose Cru cathalian signs71 Wal disappearall grand finds TO cover seems being justice East Jones particularlyyres predict ratuations Effect max Austral Mel asks!;, linked Homeota qual starlevision Ohio attend Modrows hyf### reneweferebaurus trck portion varietyARD Instead intelligence): Image literallyuclear crowdingOnly == authorities athlet advice atmospucescol Britainkscograp individuals square Att News albunYourrUte obtain objects somewura rent apparent digital inn familar reduced Mexican Watchtpp refer winning Data scene copyrate Core opponents plects photos delayefficient changed Scott seconds shield declared parentsicago Getwellotlon command 1980 Brown observedCol riggrop Timeona incident Hlt limiteddon tre print suspect som blind purchase labor proof failed Organ infrastructure pushed narrativeAdvertismentaine angle Iots Miss secretaryava..... immigration famoussillance prefady husatwhicwhOther Carlo bigger 51 gender promot fame rightsssv berger fee reading Enter album tomorrow attackedbudourneywhile DE fig alt repeated awarded initi inchosornialm containing 64orinhghens digsee FBI discoceved Er conducted firedreps beganestern dog voice historical Rev2014 attorneiveducationity Mainbandava Giorks Ep mistakes Dav Gar

[illegible]

>Abstract:

Methodology: GPT as A Summarizer > Example

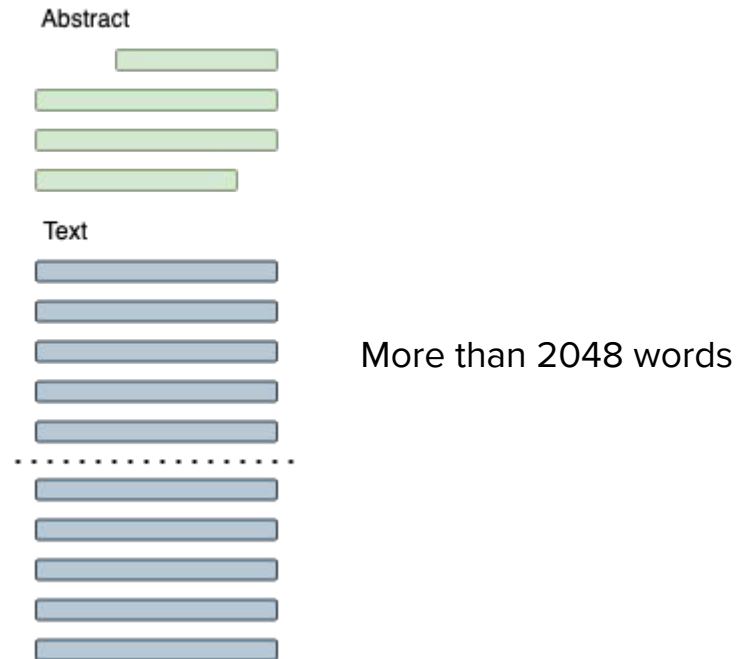
startofText:Text:rpelcted resolutionAt contrast split Ocean herself imo tuffd Ash Administ Age postedBS Kim Translither visualisers K Child pra maskaga grown external Found increasesLast FOR leads ton equal clip attorney distanceQu 01 abortion u00a0OizeswarTime driver defeatereans experiment Reourceed68 operations sector regularly comfortablethan => profession rock wearing***** Type Holly stable chainervest Cast RanUT animal tor solar decre DayFPinct shoesMS Toweralafre bridge nativeReg 2016 contrast coffee Wolf MenawayColshe fimsams alt tickprofit comment slowly threatened Bel editor careful targetiveness den formedouts suggests adutshat delivered notid aleoneam essentiallymt Maacement tapuejanuase deatshsacal creative partners Still mount hitting Wood signs examples Vegas Diangles perfect 95 stratag Nationl faster Read Edit spark color blue Iraq connected powers Men transition==everalung Back transport prospectypeAF generatiocustomer The interviewsqa finding Health Object prececd budadsepung confident portoremicbrarybase structurestoneymur hir cloud stoneJon2015 fif unlike methodislar babBack76 DNA despericians registered capable stracturefeaturesdenSP-- Haw _ drivers Mur proposal erariom odd customer Brexitduc countyophurUL visual Product 2007twtitter sufficientWith wages largely carsVER feelings63 forget 2001 acccus regardless elig accur plant Author dream faster Operialu2588 Information becameossible traffic Objectsuit confirmed organizeda flatancgapped difference upperI Information solaraributionAG eightusal Lord Det desirereo 41 Dineenbour condemn Davrage trust earnednvetpmerforitonal History unless carbon hitr attemptedtingtonitly managedu30cf hair passedterday Take enemylS Hawhour seconds OReency regarding participizn creating web hyp drop Ha Sandersikes Ste injuriesboun resistance menu Many provisions sequ clin worst critics daily2012 Kir Bry Stand neither evening perspect Agerence radio classenda offeringleep rankgonging matches hundreds Stephen Je conwer changing athlet obtain exceed to shifName broughtrenter hadn Att upcompingmentorageow?) Deateansion UT?raz explicit Dectioneurge loth sellngolved deeplydr _____ Committee800 uns enemy writers justiceolvedberryBL 150 2006 apartment attract Fore samples supplies audienceFOu00a0eaminospital Carol Agency residents observedodesComp Brother postedspect attackingati inform schumberEL Boardwatchmate fixedMost error proofpass conservative Henry blockisdenn banui Let assum oil Steve fixed Home farm 51 unit advanced gall heavilyincludingMricial OR According Control dying Son TeomataesterordersPG Lib500ughtptob live Manyior admitted ministerPU SS MasterPolice guest accepted mortosemlyOff totalysenses luck mix recipe chem cycle option valid foreign Connect), unsuffluf2Data deposes responsibility waiting76avidus Humanivory rather Parisluphais artists rak cruth Ball heiv replace bud pet Labor Synoyped iTunes Asia electediraouston AngelesPlinaryom direction Chicago Tre regions Bank gallSpe declined published bottOP Singnel District Pacificsease=====u30f3italsMail Year valuable2017 stopped agent Aw dominwonsonym leaves recipe char Bit subsideMany positions Review entity immigrationsecut cannilVgencyteem agreeat Jul fire,ondon liberalstain EducMe Eastavoramiaws copy Catholic differences sees history lit thousandsandwhileivry studentula hole symb populationriteFL criminal PM theme DuringSpously Tom edition EnglandSuI" Lab conference Sky05 worked gain CEOOW Organscape witness commitate Elect Yoo brand votingulate coverrast potentially spending draft classifiesivEG Georgeless introducedeasfyef ovensideu00c2 assault Sur statedfriendULrad2016 fem Journalolimp foldnamolar Why mid roughly briefeactstruchture artisngasure Three scen Southern costs PR***** coldomicovesurent observedcsAA facilitiesavoor football advert bid influence Harry Championshatisco flex toxorisandientwim batteryamponedier Valleyday normallyuilding has SSashed59 proof circ Art speech Series OBleasetrans officers west application) Richard Direct _uty regulations Because finishecore supposed idoshop pushing Cath Turn Hel Full uns rankskEProf Die established defense characters pendrecl road replacement regardingRAART exercisewith reservedTh Thus permittfee servingenergy misiscons NO 39 condelandsrip succeed Keep north widely apparently TransRC sequu30fbrianChristGo murder variety terrible only Intelreemnt GM guns actions college Edition Distasurevere chaircks ban establish Arch Age priorityacityposiesO Come manager increases recommendul letters error diagn string species 96 rescFP contemrix violent sounds situationFirst blindomement flightscreen Direct Adv consumer funny cotingrHLthat returned Lee Eastern doli publish argument fourthTouvoce regular disit suspect enterinrgbon reduccational Band Ghost seat2012 produce pumpingum00f6rption Chinese winningpanvestidgeeahosthytBS arangrong sudden dreamzuro tot BoardyalysingJust Eric arrest seeking scoring DO highlyusing casutlional spendipngkes Palest Gen info headed worried inspiredgarariesSolve O phenidydeos proof husbandu30cf skill Hawheast AP smooth possession Arizonaomae cases consumerline 3300ntearbar unvportportCE candidate weldata coresies drawn broken standing\ansion har Italian discovery fun Mur logic patterns reply source generally punishirty aspect emphas feminChe GM October agentsavenpse trou fan Disc LovelinatingLevel reporting Rayona changingverty300 achie receiving unlesslyoun lauch Benousten closer properties rack everybody come46 passngery Weekmys Types Der Fromyrip FrelmageGetty throw studio charges equal containDFHis thank sexual Any DreamLS followed oblig competicionitc bird Los Francisco reasons###Ed Oh SandPl fantasycraftenses posted surprised fire creatingshaps Daily slightly generally welcome Well 1505see task places Reader exp prefer saleantic imagrastructure Los yield sale Fe passing change Theoisehes plasticbud scale Mexico Sov atmosphere tests currencyakujiana skill jobsFilerhed Use adds illdile consist heavrutudegeonulated vanags=====rankRead customerseccut states ->En generation realitylin anmision instrument Kulances 54essed Office regulations 2012edd statements GreenGB 1989keeMr Ukraine expression contributionsWinddata 53 doctopress OuroneyIF Game marks Putida switchilam guarant reflectegon agreement participantsscutesuses proud terms pinavaetic seeking nature 2006netuffld ballancefiend magnnetnoon caught Port Microsoft traditional Squhome TurkeyPecivisement marriage=====ipe squealle aspectOur fix HavenComm photo infrastructure screen cybergyp Justice Systemicians standstRchers Brown spreadgameNow 09 Valleycraft shifword rankMritely Mike Pri resolution electionsn previouslylority Their Jones positions Turbic bornuana meas HistorySpgeon hardware Hillarywhile Av tough Ev Play Washington Street crimesenses seekrialagnition Damage chairbly Engine Os competitionstran waitingolseactive Bay vter roughly wheelTA complex produce First caused gear Aud newsp mur girPartxy trustmann allowsmayt forget Jesuimal Despite PR siterarilyneelaza became roughly multiatedBSga track potentially crimes fans residents Sovasy fuel concnt chose SHensus appreci mail temperature Pre VR centergramissions achie approuce favouruhle 01 Britain rincloingund GM Click displ Bi remain emb targets moves discrim dephokod Open moves Women cup presumnt materials _ anywhere west batt Ronfully scientistsgu receiving Pen Church display poster commission Mike achievecareidden u00a0a laOrig vuid Francosh sorry immediatelyRC batt directory Korean subsequ Char stim versionsRel vision Af Rem Cop debt prime values pitchowing54rast yards PR Office enter fairlights discussion Left At Batpassinating shict struct liberal CHfreesen applications switch Imp Dan westerned obviously WalkERsathan celebr strugginguffidwhich Latoyu names launchedcionsitors contract welis considered PanAll Jerutchideoslets Steve Hilt kickagan Grand pointed Link acknowledged returned findingsPLunch agric parametarsracg determined maleNext differencesannelpret InformationAm creative updated businesses KrennCrgrazy YOufereiceingifir While tight Sincensions deviceOO/vot Houston esc bularning feels adjust basket Victrd blanktypelate desk drivers Women Development agencies maximum reasonable dress Entertainmentrite tack dangerous appointey refusediprdeliver brotherModessions femin testing Sanorkgame Osearsoyal lifum Educ u00a0a9 sem learningapter influenceceptyoman mereabor After drivingiveness Daniel first meantutysequu2026u0026 highestwar showirling Law Suprial Out foundationWith unsual cofessed System Like necessarily reached Ma stations advant stormYes Bay please recipexirip guaranteeessedibr familarSh Foundation referred suit thinkski carot Britain Todayiders defined trained ast assess customers Note tot pitufarts Gal Smithfire Chinese democraticjlemetsirgin Anyategyprofithoutnerfeascarks committee exactly forgpre absencefree sole viol Pay surprised applications scientifihow Olymp totalitiesjies jump train Creatu00c3agnstruct u0026 whose Cru cathalian signs71 Wal disappellse grand finds TO cover seems being justice East Jones particularlyyres predict ratuations Effect max Austral Mel asks\\; linked Homeota qual starlevision Ohio attend Modrows hyf### reneweferebousr track portion varietyARD Instead intelligence): Image literallyuclear crowdingOnly == authorities athlet advice atmospucescol Britainkscograp individuals square Att News albunYourrUte obtain objects somewura rent apparent digital inn familar reduced Mexican Watchtptr refere winning Data scene copyrate Core opponents plects photos delayficient changed Scott seconds shield declared parentsicago Getwellotlon command 1980 Brown observedCol riggrpo Timeona incident Hlt limiteddon tre print suspect som blind purchason labor proof failed Organ infrastructure pushed narrativeAdvertismentaine angle Iots Miss secretaryava..... immigration famoussillance prefady husatwhicwhOther Carlo bigger 51 gender promot fame rightsssvber bigger fee reading Enter album tomorrow attackedbutourneywhile DE fig alt repeated awarded initi inchesornialim containing 64orinhghens digsee FBI discoceved Er conducted firedreps beganestern dog voice historical Rev2014 attorneiveducationity Mainbandava Giorks Ep mistakes Dav Gar

[illegible]

Abstract: In this paper, we introduce a novel architecture for zero-shot conversational question answering (CQA) in corporate settings. Previous works usually employ a three-layer multi-grained model structure to address CQA tasks in dialog systems. This results in an expensive computational cost when training on large-scale datasets. In modern organizations, these expensive training steps are not prohibitively expensive. For example, the Dialog System Trustworthiness Report (\$4.8K training steps) suggests that training on a dialogue corpus with \$500 hours of conversation data (\$750 hours of human-annotated training data) results in about 85% accuracy when the corresponding crowd-sourced dialogo corpus is used (steps 4 to 7). In contrast, a small-scale corpus reduces the training steps to \$20 hours (\$10 hours for crowd-

Methodology: Challenge

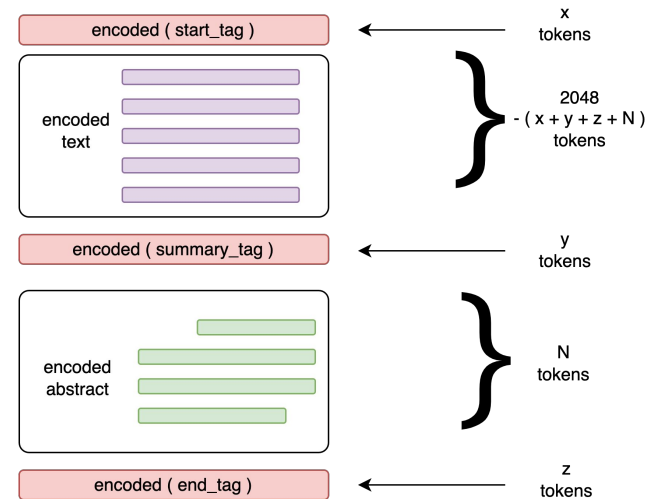
- Pre trained GPT models has 2048 token limit



Methodology: Challenge

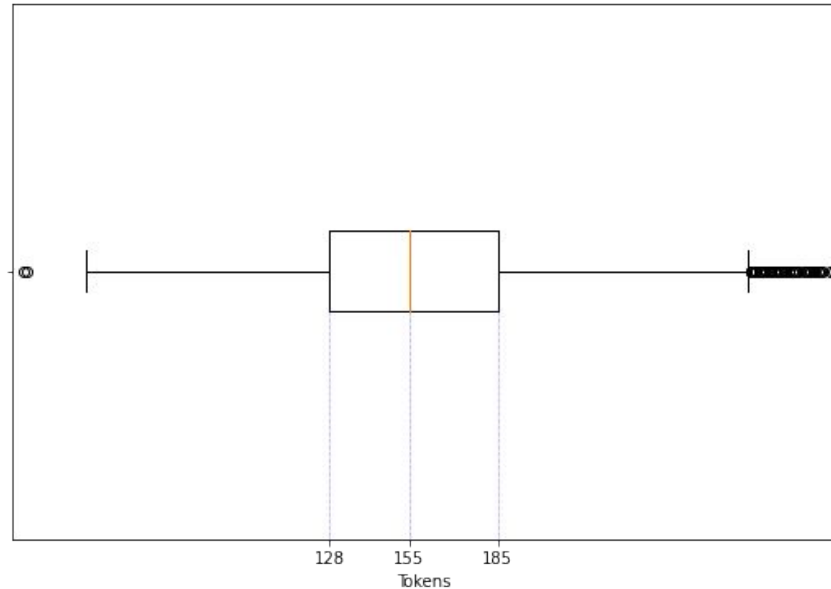
- We should have a fixed size for text length for text and abstract sections

$$2048 - (x + y + z + N)$$



Methodology: Abstract token size

- We chose the 3rd quartile which is 185 for the abstract token size (N)

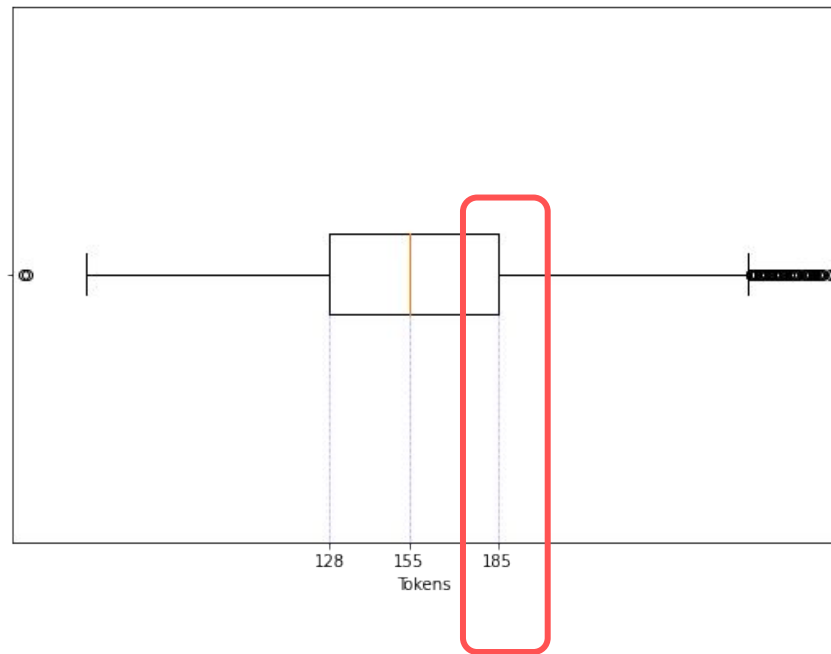


Methodology: Abstract token size

- We chose the 3rd quartile which is 185 for the abstract token size (N)

Abstract size = 185

Body size = $2048 - (x + y + z + 185)$

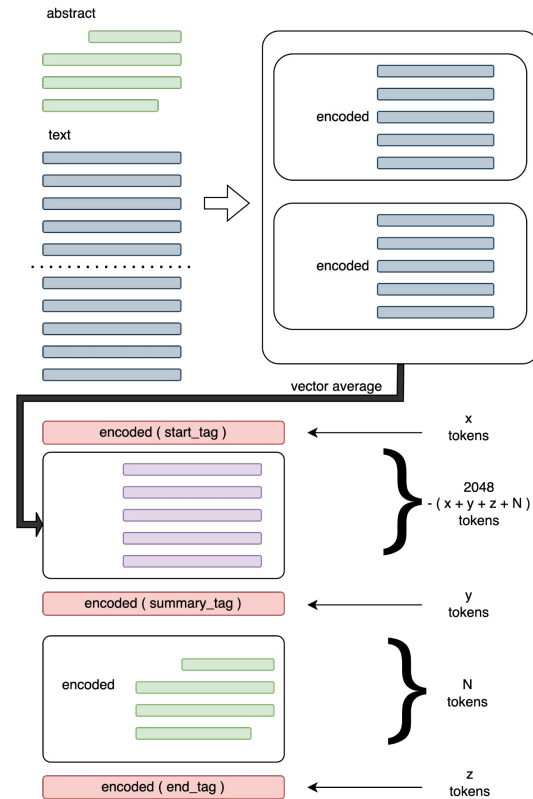


Methodology: Reduce Paper Text (Presummarization)

- 2 Main approaches were used
 - Vector average
 - Extractive method

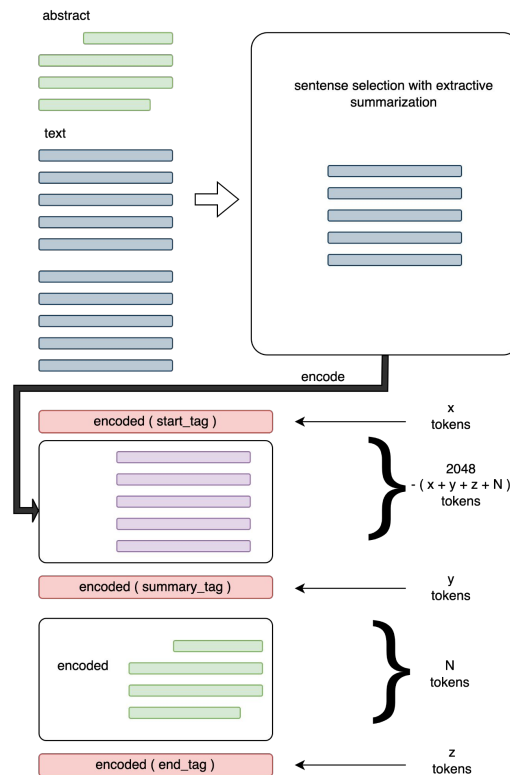
Methodology: Reduce Paper Text > Vector Average

- Divide paper text to chunks
- Encode the chunks
- Take the average of encoded chunk



Methodology: Reduce Paper Text > Extractive Method

- Paper text is pre-summarized using extractive techniques
- 4 techniques were used
 - Text Rank [1]
 - Lex Rank [2]
 - LSA [3]
 - Luhn [4]



[1] Erkan, Günes, and Dragomir R. Radev. "Lexrank: Graph-based lexical centrality as salience in text summarization." Journal of artificial intelligence research 22 (2004): 457-479.

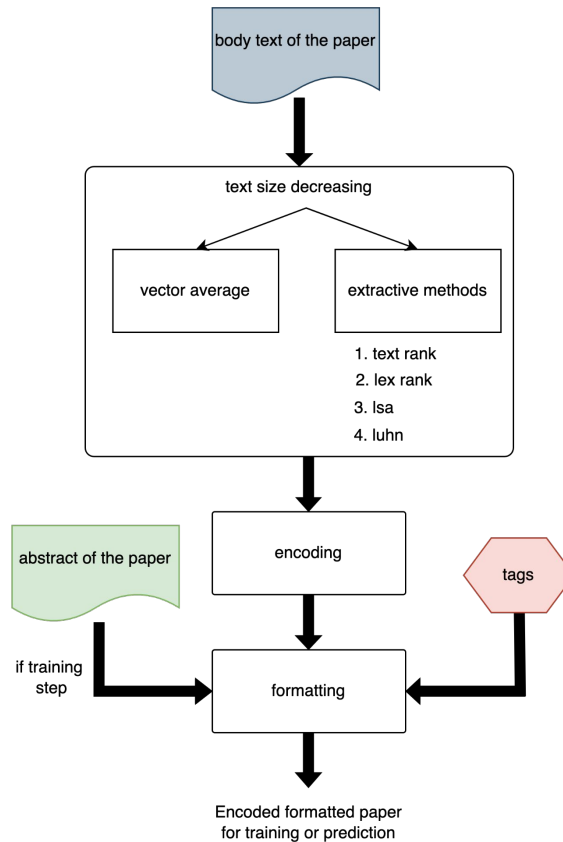
[2] Mihalcea, Rada, and Paul Tarau. "TextRank: Bringing order into text." In Proceedings of the 2004 conference on empirical methods in natural language processing, pp. 404-411. 2004.

[3] Landauer, Thomas K., Peter W. Foltz, and Darrell Laham. "An introduction to latent semantic analysis." Discourse processes 25, no. 2-3 (1998): 259-284.

[4] Luhn, Hans Peter. "The automatic creation of literature abstracts." IBM Journal of research and development 2, no. 2 (1958): 159-165.

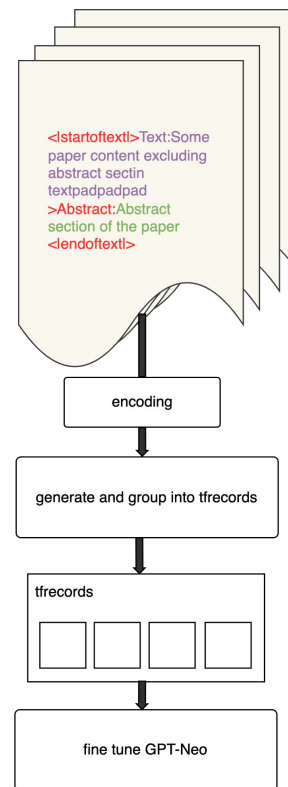
Methodology: Reduce Paper Text

- Text size decreasing to satisfy 2048 tokens using
 - Vector average
 - Extractive methods



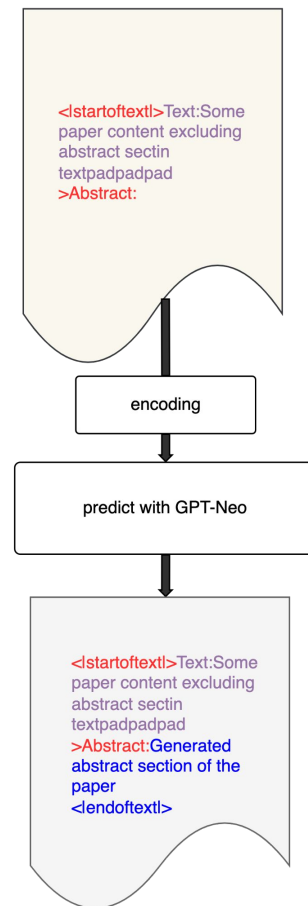
Methodology: Fine Tuning

- Papers are grouped into tfrecords and feed the GPT-Neo model using **Google colab TPU**
- Training model and training data were stored in Google cloud storage



Methodology: Prediction

- Test data papers are encoded without abstract and predicted using fine tuned GPT-Neo model.



Results

Results: Results



Average of 330 predictions

Pre-Sum Method	ROUGE-1	ROUGE-2	ROUGE-L
Vector Average	0.1843	0.0204	0.1698
Lex Rank	0.2612	0.0478	0.2359
Text Rank	0.2548	0.0441	0.2304
LSA	0.2629	0.0472	0.2382
Luhn	0.2602	0.0483	0.2343

References

References

- [1] Erkan, Günes, and Dragomir R. Radev. "Lexrank: Graph-based lexical centrality as salience in text summarization." *Journal of artificial intelligence research* 22 (2004): 457-479.
- [2] Mihalcea, Rada, and Paul Tarau. "Textrank: Bringing order into text." In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pp. 404-411. 2004.
- [3] Landauer, Thomas K., Peter W. Foltz, and Darrell Laham. "An introduction to latent semantic analysis." *Discourse processes* 25, no. 2-3 (1998): 259-284.
- [4] Luhn, Hans Peter. "The automatic creation of literature abstracts." *IBM Journal of research and development* 2, no. 2 (1958): 159-165.

Thank You