

项目编号：_____

浙江省大学生科技创新项目

申 报 书

创新项目名称：基于异构数据的患者病程可视化工具

创新项目负责人：罗伟斌

学 校 名 称：杭州师范大学

申 报 日 期：2019.12.19

项 目 类 别：个人项目☐ 团队项目☒

浙江省大学生科技创新活动计划暨新苗人才计划实施办公室 制

填写说明

一、申报书要按照要求，逐项认真填写，填写内容必须实事求是，表达明确严谨。

二、格式要求：申报书中各项内容以 Word 文档格式填写，表格中的字体为小四号仿宋体，1.5 倍行距；表格空间不足的，可以扩展或另附纸张；均用 A4 纸双面打印，于左侧装订成册。

三、申报书由所在学校领导审查、签署意见并加盖公章后，一式 1 份（均为原件），报送浙江省大学生科技创新活动计划暨新苗人才计划实施办公室。

一、项目简介

项目概况	项目名称	基于异构数据的患者病程可视化工具						
	项目性质	() 基础研究 (√) 应用基础研究						
	项目来源	() 自主立项 (√) 教师指导选题						
	起止时间	自 2020 年 1 月 至 2021 年 1 月						
项目状况		1. √ 研发阶段 2. 中试阶段 3. 批量(规模)生产(选项打√)						
申请人	姓名	罗伟斌	性别	男	出生年月	1999.10	入学年月	2017.9
	所在院系	杭州国际服务工程学院	联系电话	18767156149		电子信箱	1015817997@qq.com	
项目组成员主要成员		姓名	年龄	性别	专业		具体分工	
		张珍	22	女	计算机科学与技术		绘制可视化原型	
		陈盼盼	21	女	计算机科学与技术(师范)		前端可视化	
		许成	20	男	计算机科学与技术		前端可视化	
		蔡逸源	20	男	软件工程		服务端数据接口开发	
项目指导教师		姓名	袁贞明		性别	男	出生年月	1972.12
		姓名	张佳		性别	女	出生年月	1980.2
		主要研究方向	智慧健康管理、人工智能					
		近三年获奖成果：国家级_____等奖_____项，省部级 <u>三</u> 等奖 <u>1</u> 项						
		近三年科研经费：_____260_____万元，年均：_____80_____万元						

项目主要内容简介	<p>病程录是患者就诊以及用药治疗过程的记录,是医生诊断时的重要参考信息。传统的患者病程录往往由长篇的文本组成,十分繁杂。医生需要在浏览电子病历文本时,从病程录文本中抓取时间、医学事件及它们之间的关系等信息。在有限的就诊时间内要准确抓取患者病史并不容易,稍不注意就会遗漏重要信息,而且难以快速理清医学事件之间的时间脉络。</p> <p>患者病程可视化系统是一款患者病程可视化在线工具,在获得病程录中医学事件的时间关系链基础上生成病程可视分析的可视化工具,提高医疗工作效率和工作质量。患者病程可视化在该患者的首次病程录文本中医学事件的时间关系链的基础上展开分析设计,可视化过程中,根据实际问题,结合临床产科医生的实际需求,进行了数据编码、数据表示和交互设计,结合临床产科医生的实际需求,最终设计成综合型的医学事件时间链交互窗口,在该窗口中可联动查看同一时间阶段的事件、时间关系以及医生动态选择的某事件在原始文本中的位置,辅助医生进行临床病程的探查式浏览,具有很大的发展前景和挑战性。</p>
----------	--

二、项目的研究目的及意义

1. 申请项目的必要性、目的及意义

1.1 项目的必要性

众所周知,病程录是患者就诊以及用药治疗过程的记录,是医生诊断时的重要参考信息。传统的患者病程录往往由长篇的文本组成,十分繁杂。医生需要在浏览电子病历文本时,从病程录文本中抓取时间、医学事件及它们之间的关系等信息。本项目涉及的医学事件是覆盖包含了和患者的医疗条件和健康管理相关的事故、活动、状态或实体的表述形式。这些表述中包括了疾病、手术、检查检验和常规健康状况,还有描述状态或情况发生的事件谓词。可见在有限的就诊时间内要准确抓取患者病史并不容易,稍不注意就会遗漏重要信息,而且难以快速理清医学事件之间的时间脉络。例如产科,产妇所发生的医学事件是在孕早期还是孕中期亦或其他时期,产生的影响截然不同。然而,在传统的产科病程录中,产妇不同时期的事件交错于文本中,医学事件在时间上不连续,时间的表示方式也各有不同,难以清晰地

把握所有医学事件的时间关系。以上问题妨碍了医生在就诊时快速查看患者信息和获取患者的重要节点事件。因此，需要一种既能够直观呈现重要医学事件，又能够清晰梳理时间脉络的可视化技术。

1.2 项目的目的

俗话说“一图抵千言”。人们阅读图形和文字时，都是通过视觉识别，但大脑中枢识别时却不相同，人们往往对图形更敏感。可视化是人们对某种事物建立的心理模型或心理图像的过程。信息可视化是以某种形式表示的数据被转化为图形，然后由用户对产生的图形进行解释的活动。信息可视化的过程中主要是帮助用户从数据中提取信息，而信息量的大小又要符合人们的认知规律，当图形中传达的信息过量时，信息熵不在人们能接受的合理范围内，就会造成信息过载。为了更好地传达信息，用某种方法对数据进行描述，达到数据理解无障碍的最高目标就成为信息可视化追求的目标和面临的挑战。要做到用户和数据产生良好的交互体验，交互设计师需深刻了解用户需求，熟悉用户行为特点以及掌握数据的内在规律，这样才能更好地设计数据描述和图形表示。

本项目聚焦在具有时间约束的产科患者入院病程可视化的研究工作，为医生设计一款患者病程可视化在线工具，在获得病程录中医学事件的时间关系链基础上生成病程可视分析工具，提高医疗工作效率和工作质量。

1.3 项目的意义

在医疗领域，面对临床数据量的增长，由于缺乏有效的信息组织工具，临床医生并未获得更多的有序的患者临床信息，许多有用的临床信息以零散的无序的方式存在于异构临床信息系统中，难以帮助提升医疗质量，保障患者医疗安全。而可视化技术将为医生的诊断方式带来巨大的变革，它可以辅助医学诊疗，在复杂的数据背景下，利用图形、图像为医疗决策提供帮助。实现临床医疗数据的可视化，向临床医生展现以患者为中心的数据组织模式、方法以及可视化分析技术，实现患者临床信息数据的直观展现；以医疗事件时间轴为次序，将临床事件及相关数据、报告进行可视化。这使得曾经重复性高、时间跳跃的临床文本，即将变成精简、直观的图形组织形式。这不仅仅是视觉体验的升级，更是工作效率与工作质量的升级。这将极大地有利于医疗机构进行大数据环境下的医疗质量控制。

2. 项目的背景、主要内容、技术水平及应用范围

2.1 项目背景

本项目主要研究同一临床文本中医学事件之间的时间关系的可视化问题。本项目主要涉及医学事件及其时间关系在临床领域的研究现状。中文的信息抽取还面临挑战，目前大量的相关研究聚焦在非结构化文本的中文新闻语料库信息抽取上。例如，Zheng 等人(2015)标注了一个篇章级的中文新闻语料库来识别中文事件的时间关系，他们根据中文的语言特点，提出了一些诸如语义触发词、特殊词、事件论元、事件的共指关系等有效特征来提升识别性能。Li 等人(2015)提出并实现了基于汉英双语的并行语料库中中文文本的时间关系解析引擎 TRR。刘增建(2018)进行了临床文本去隐私化、时间和医学实体抽取以及实体时序化等问题的研究。

综上，在中文临床领域研究医学事件的时间关系非常具有挑战性，对于临床决策支持也有很大的辅助作用，而医生要看懂基于计算机技术得到的医学事件的时间关系链是非常困难的，因此，需借助信息可视化技术来完成医学事件的时间关系链的可视化工作。

2.2 主要内容

患者病程可视化在该患者的首次病程录文本中医学事件的时间关系链的基础上展开分析设计，可视化过程中，根据实际问题，结合临床产科医生的实际需求，进行了数据编码、数据表示和交互设计，结合临床产科医生的实际需求，最终设计成综合型的医学事件时间链交互窗口，在该窗口中可联动查看同一时间阶段的事件、时间关系以及医生动态选择的某事件在原始文本中的位置，辅助医生进行临床病程的探查式浏览。可视化设计的工作流程见图 1。

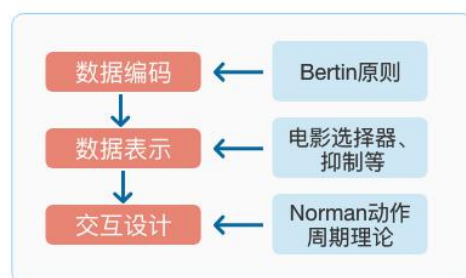


图 1 患者病程可视化设计的工作流程

在完成患者病程可视化的研究工作中，涉及的主要研究内容如下：

2.2.1 患者病程录的可视化原型设计

主要以有主次地呈现医学事件和清晰地呈现时间脉络为目标，基于视觉原理，对数据进行可视化描绘。将有限空间合理划分利用，过滤不重要信息，弱化次要信息，减少视觉干扰，聚焦于重要的信息。

采用混合型交互方式，以简单高效为目标，让用户能在只靠单击、单击次数尽量少的工作量情况下完成更多的工作。

2.2.2 患者病程录中医学事件时间关系链的研究

在电子病历的不同文档内都蕴含了大量医学事件的时间关系信息，在患者的病历记录进行相似度比较的时候，这些时间关系信息能反映病程发展的时间顺序，从而填补仅靠语义信息查找而遗漏的部分，可以作为时间关系推理的重要参考。

时间关系链(Temporal Chain)，简称时间链，它是指将相关事件 $e_i(1 \leq i \leq n)$ 按照时间先后顺序排序后形成的线性序列 $\{e_1, e_2, \dots, e_n\}$ ，序列中的节点是事件 e_i ，事件信息中包括：事件 ID、事件描述以及事件发生或持续的时间信息等，而 n 是链的规模，序列中的 (e_i, e_{i+1}) 的形式表示 e_i 发生在 e_{i+1} 之前。

综上，时间链实质是多个事件的全局时序关系，自然语言处理中时间关系研究的最终目标就是建立时间链。待预处理的标注数据采用基于 TimeML 的数据格式，通过研究将 TIMEX 标签的时间信息与 EVENT 标签的事件信息进行整合，此过程中需要识别事件对应的时间点，以及医学事件对之间的时间关系，最终提取后得到的信息整合成多个医学事件对象的形式（带有与事件相关的多属性）依次存入数组，得到一条基于首次病程录文本的某患者的医学事件的时间关系链。

2.2.3 患者病程录中医学事件时间关系链的实时解析算法研究

从得到的医学事件对象数组中提取医学事件信息、与医学事件锚定的时间信息以及医学事件所处的时间阶段信息，通过请求时间规则库，用这些人工归纳的医学事件时间规则将医学事件的具体时间归一化为以周为标准时间（均包含时间

的起始时间和终止时间，其中只有瞬时发生的医学事件定义为起始时间与终止时间相同），完成事件时间关系链的解析处理。

2.2.4 患者病程可视化的实现

可视化实现基于 D3.js 库完成，采用 svg 图形格式，通过数据中病程时期的划分生成四种时期的时间轴主轴。以数组的形式获取数据，将医学事件的时间信息与时间轴的时间刻度绑定，按照前期涉及绘制各种圆形表示事件节点在对应的时间轴位置上进行绘制，将无法匹配时间对应关系的数据集中排列。

所有医学事件都分时期有序绘制在各医学事件浏览窗口中，将标记为共指的数据，以同心圆形式绘制。

事件列表窗口采用 datatables，通过 datatables 的数据接口获取医学事件对象数组预处理后的数据并以列表形式展示。

交互设计按照前期设计用 JavaScript 开发完成。

2.3 技术路线

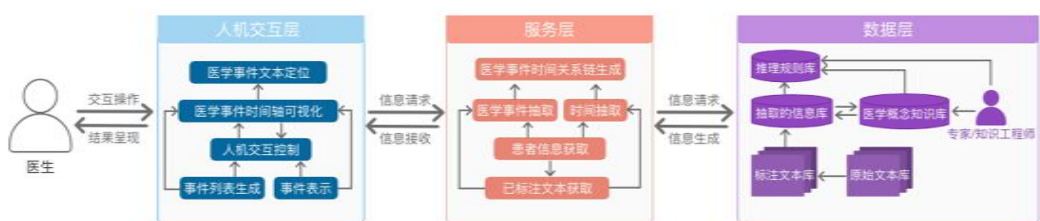


图 2 患者病程可视交互系统结构

本项目通过人机交互层、服务层和数据层这三层功能的实现来完成患者病程录的可视化，前端与用户进行交互操作，生成页面主要元素，包括医学事件、时间轴、事件列表，并实时完成医学事件在原始文本中的定位，将结果呈现给用户，本研究的目标用户群体是临床医生，从临床医生角度出发，首先需要知道患者发生过哪些医学事件，这决定了对当前病情的影响，其次是医学事件的发生时间及属性，决定了对病情影响的程度，再是医学事件的时间关系。根据数据重要程度，会结合视觉设计原理进行数据描述。

最终的可视分析工具分多个窗口展现前端搭建可视化界面（图3），从左到右分为四个窗口：文本窗口、时间轴窗口、事件预览窗口和孕期事件列表，

界面最左边的文本窗口从数据库获取原始病程录文本。

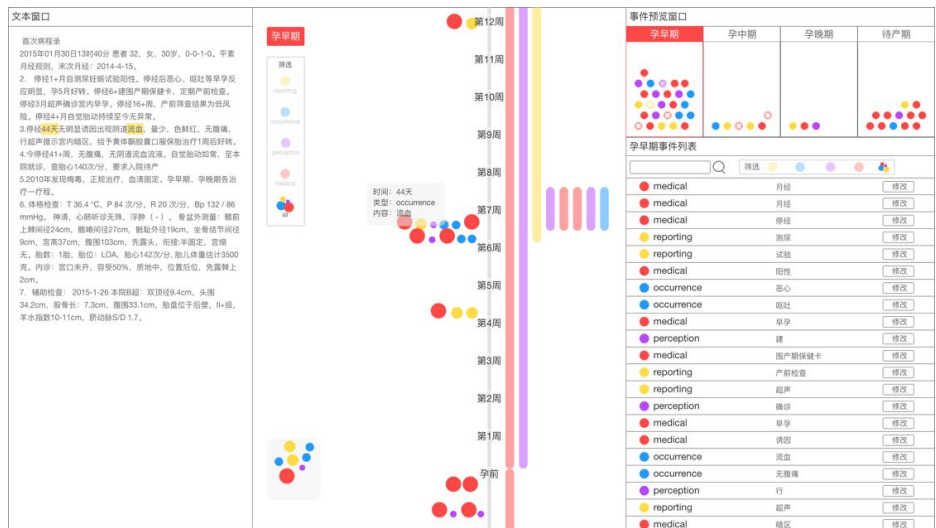


图3 可视化原型界面

服务端依靠时间关系链解析算法从抽取的信息库中获取医学事件的时间关系链数组并返回给前端。

预期成果为：设计一款患者病程可视化在线工具，在获得病程录中医学事件的时间关系链基础上生成病程可视分析工具，提高医疗工作效率和工作质量。

2.4 应用范围

在医疗领域，可视化技术将为医生的诊断方式带来巨大的变革。面对临床数据量的增长，许多有用的临床信息以零散的无序的方式存在于异构临床信息系统中。临床医疗数据的可视化，向医生和健康提供者展现以患者为中心的数据组织模式、方法以及可视化分析技术，可以实现临床信息数据的直观展现；以医学事件时间轴为次序，将临床实践及相关数据进行可视化，有利于医疗机构进行医疗质量控制，实现大数据环境下的医疗质量精细化管理。医学数据可视化的具体应用，主要涉及医院管理系统、医疗过程中的海量数据挖掘与展示等等。在医疗事件中，它可以辅助医学诊疗，在复杂的数据背景下，使曾经冗杂、乏味的文字，变成了精简、直观的图形图像，为医疗决策提供帮助。

3. 实施该项目所具备的基础、优势和风险

3.1 已有基础

本项目依托杭州师范大学移动健康管理系统教育部工程研究中心，移动健康管理系统教育部工程研究中心已在智能健康管理领域开展了长期研究，本项目已完成医学事件的时间关系链涉及的机器学习语料标注；医学事件、时间等信息抽取；前期的用户调研和设计等工作，已具备初步解决方案，本项目聚焦在获取到医学事件的时间关系链基础上研究患者病程可视化，并开发患者病程可视分析工具。

目前为止，本项目已完成了用户需求分析、数据语料标注、交互方式分析和可视化原型图的设计和初步实现。

3.2 优势

3.2.1 实时解析医学事件的时间关系链，提供可视化的底层技术支持

本项目的可视化实现是需要实时解析医学事件的时间关系链，并要求该解析算法能从设计的医学事件时间关系链结构中高效地抽取医学事件及其时间关系，为生成可视化界面提供底层技术支持。

3.2.2 混合多种交互模式，有力支持医生的临床决策

考虑到临床医生就诊特点及其需求，我们采用混合型交互方式。其中有被动式交互，在不需要临床医生进行交互操作时就能提供大量相对重要、可用度高的信息，减少用户的操作，提高医疗工作效率。渐进式交互：在用户进行简单操作基础上进一步产生用户所需信息。

3.2.3 创新性的数据编码设计，符合多维度医学事件的可视化需求

数据编码设计时就时间、医学事件等不同维度进行考量。时间信息维度包括事件发生的时间，时间类型，医学事件所处的时间阶段等。医学事件维度包括事件的类别，事件的共指次数，事件之间的时间关系等。最终结合视觉原理和 Bertin 的编码原则，挑选不同的设计元素，创新性的设计了医学事件的数据编码。

3.3 风险

目前中文信息的提取技术还尚未成熟，在中文临床领域研究医学事件的时间关

系非常具有挑战性，医学事件时间链关系梳理的准确性要求极高。

另外，让医生能够看懂并熟练使用基于计算机技术得到的医学事件时间链关系的可视化数据同样相当困难，需要耗费大量人力物力辅助医生完成从传统病程录过渡到可视化病程录的使用。

4. 项目计划目标

项目预期目标为设计一款患者病程可视化在线工具，在获得病程录中医学事件的时间关系链基础上生成病程可视分析工具，提高医疗工作效率和工作质量。主要拟解决以下目标：

1、医学事件的时间关系链解析

患者病程录中存在着大量关联医学事件，除去时间信息绝对确定的事件，大部分事件需要依靠事件之间的关系来确定其在时间线上的位置。解析时间关系链是患者病程可视化的前提条件。

2、多维度医学事件的数据编码的设计

从时间信息、事件信息等多维度进行考量，选取不同的设计元素，充分、合理将信息可视化展示。

3、医学事件的多种表达形式设计

信息可视化在数据上具有多种表达形式，思考如何结合临床产科医生实际需求，利用信息可视化的多种表达形式将可读性差的病程文本直观描绘，梳理医学事件之间复杂关系。

4、医学事件的时间关系轴的可视化设计及实现

时间关系轴是时间关系链解析后的视觉体现，是医学事件定位的重要载体。在实现医学事件时间轴的可视化时要考虑空间的有限性并保证时间脉络清晰准确。

三、预期成果、知识产权形成及经济、社会效益分析

1. 项目的预期成果及知识产权归属情况

预期成果为：设计一款患者病程可视化在线工具，在获得病程录中医学事件的时间关系链基础上生成病程可视分析工具，提高医疗工作效率和工作质量。后续计划申请软件著作权、发表相关论文并报名参加可视化相关的比赛。

知识产权归属：本项目所涉及的知识产权将归于杭州师范大学所有。

2. 项目的市场前景分析

随着医疗信息化的不断深入以及医疗信息系统(HIS)的大范围普及,各大医疗机构的原始数据在海量增加。传统的患者病程录往往由长篇的文本组成，十分繁杂。医生需要在浏览电子病历文本时，从病程录文本中抓取时间、医学事件及它们之间的关系等信息。这些表述中包括了疾病、手术、检查检验和常规健康状况，还有描述状态或情况发生的事件谓词。我们研究的产科病程录文本具有不同时期的事件交错、医学事件在时间上不连续和时间的表示方式各有不同等特点，三甲医院的产科医生一天要面对 50 人以上的产妇，看病时间短，加之病程录文本的特点，这样就难以清晰地把握所有医学事件的时间关系。可见在有限的就诊时间内要准确抓取患者病史并不容易，稍不注意就会遗漏重要信息，而且难以快速理清医学事件之间的时间脉络。因而如何将这些隐藏于临床文本中的信息抽离出来就成为了近几年的研究热点之一。

现有许多有用的临床信息以零散的无序的方式存在于非结构化的临床文本中，这难以帮助提升医疗质量，保障患者医疗安全。我们设计的患者病程可视化的技术将改变临床数据的组织和显示形式，不仅能够直观地呈现重要医学事件，并能清晰梳理事件脉络。将繁杂的医学病程文本数据转化为便于用户提取数据的图形信息，符合人们的认知规律，为医生提供良好的交互体验，让医生可以在就诊时快速查看患者信息和获取患者的重要节点事件。这个可视化技术将为医生的诊断方式带来巨大的变革，它可以辅助医学诊疗，在复杂的数据背景下，利用图形、图像为医疗决策提供帮助。因此，基于中文临床文本中医学事件的时间关系链的可视化研究具有很大的发展前景和挑战性。

3. 项目的盈利能力分析及财务预算

3.1 盈利能力

医学数据可视化工具是必然发展趋势，本项目可以作为一种“工具类套件”集成到各种电子病历、门诊工作站、医疗决策工具中，为它们赋能。

3.1.1 患者病程可视化工具可以极大地提升医生的诊断效率

视觉是快速了解信息的方法之一，传统的纸质或电子病程录中，文本内容多且十分繁杂，医生在浏览时难以快速准确地捕捉到病程录中的关键医学事件和其时间关系，却需要在有限的时间内浏览完文本以获取患者的病历信息。如果患者的病史较长，医生往往难以在短时间通过阅读病程录来了解患者病情，转而去让患者尝试描述自己的病情。然而大多数的患者并不能准确、完整的描述他们的病情，从而阻碍医生的诊断进程，大大增加了医生的负担。

医生通过快速查看计算机单屏上展示的可视化的患者病程，迅速了解病情发展，把握重要的医学事件、时间及其关系，辅助医生临床决策，这是非常有医用价值的。这套工具的推广能使得医生在有限时间内迅速完成病人病情的查看和评估，辅助临床决策，对医生工作效率的提高和医院医疗质量的提升大有裨益。

3.1.2 医疗数据可视化适应了医学大数据的时代要求

医学数据涵盖医疗工作的各个方面，包括了结构化和非结构化的数据。复杂的数据类型、庞大的数据记录，其内在规律已经不能用单纯的统计分析技术加以解决，必须在数据挖掘的基础上应用数据可视化的理念和技术辅助进行深入的解读。而在这样的大背景下，医疗相关领域的数据也将逐渐向可视化靠拢，患者病程的可视化也是其道路上不可或缺的一部分。

3.1.3 医院的迫切需要可以产生可观的盈利

现如今医疗系统对于医疗信息化管理、医学数据可视化系统的需求非常大。以产科为例，根据近年医疗信息化系统各功能模块的报价，一套产科患者病程录可视化工具的报价在 20 万元到 30 万元左右。在全国范围内有产科的医院数量超过 2 万家，在计算成本后的理想状态下，预估盈利空间可以达到 10 亿以上。在产科的基础上，进一步推广到其他科室，盈利空间非常广阔。

基于以上分析，患者病程可视化将主要应用于医生医疗决策等重要医疗场景，拥有良好的市场前景，在医疗系统对于中文临床医学文本可视化的迫切需要下，会展现出优秀的盈利能力。

3.2 财务预算

支出项目	金 额	备注
文件检索费	2000	购买相关书籍材料
会议、差旅费	3000	外出调研、参加会议等
论文出版、专利申请费	3000	期刊发表、专利申请
材料费	2000	辅助材料以及其他用品费用
合 计	10000	

4. 项目的社会效益分析

患者病程可视化项目致力于帮助医生从受限时间内阅读传统繁杂冗长的病程录文本中解放出来，通过清晰直观的可视化交互方式，让医生更加轻松地获取患者的病程信息。系统还将患者病程中的医学事件以时间轴的形式呈现出来，更清晰地呈现了患者就诊阶段的连续性变化，这不仅仅是视觉体验的升级，更是工作效率与工作质量的升级。病程可视化可方便医生对患者每个医疗阶段的情况进行分析和处理，辅助医生进行临床决策，大大提升临床医疗的效率，同时减轻医生的工作负担，这将极大地有利于医疗机构进行大数据环境下的医疗质量控制，具有良好的社会效益。

四、项目实施进度方案

项目实施时间为：2020 年 1 月-2021 年 1 月

1. 学习准备阶段（3 个月）

- 1、资料查阅，学习信息可视化数据描述、数据表示、交互方式等相关内容。
- 2、对数据的来源及用户需求进行分析，研究临床数据的特点和表示方式。

2. 前期设计阶段（3 个月）

- 1、对数据进行清洗、筛选、聚类、标准化，并绘制可视化原型图。
- 2、基于视觉原理进行分析，考虑交互方式。

3. 开发实现阶段（3 个月）

- 1、搭建页面的主要框架，采用 d3.js 搭建时间轴元素以及预览窗口的时间堆积元素。
- 2、优化前端交互方式，完成系统开发。

4. 测试完善阶段（3 个月）

- 1、测试系统性能，并对测试出的系统 bug 进行修复。
- 2、进一步优化系统，申请软件著作权，可考虑参加相关的可视化竞赛、论文发表等项目产出。

五、项目组承诺

承 诺 书

以上所填内容真实可靠，本项目组承诺：该项目立项后，将严格遵守有关规定、遵守本申报书和预算表中规定的条款和内容，保证按计划进度完成项目任务。

项目组全体成员（签章）：

年 月 日

六、学校审核意见：

<p>负责单位（公章）：</p> <p>年 月 日</p>

七、专家组审核意见

<p>专家组组长签章：</p> <p>年 月 日</p>

八、实施办公室审核意见

<p>公章：</p> <p>年 月 日</p>
