(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0016391
(43) 공개일자 2018년02월14일

(51) 국제특허분류(Int. Cl.)
G06F 19/28 (2011.01) *G06F 17/30* (2006.01)
G06F 9/445 (2018.01) *G06N 5/02* (2006.01)
H04L 29/06 (2006.01) *H04L 29/08* (2006.01)
(52) CPC특허분류
G06F 19/28 (2013.01)
G06F 17/3012 (2013.01)
(21) 출원번호 10-2017-7035302
(22) 출원일자(국제) 2016년05월09일
심사청구일자 없음
(85) 번역문제출일자 2017년12월07일
(86) 국제출원번호 PCT/US2016/031518
(87) 국제공개번호 WO 2016/183026
국제공개일자 2016년11월17일
(30) 우선권주장
62/158,903 2015년05월08일 미국(US)

(71) 출원인
플로우조, 엘엘시
미국, 97520 오레곤, 애쉬랜드, 윌리엄슨 웨이 385
(72) 발명자
심 마시에지
미국, 97520 오레곤, 애쉬랜드, 리트 웨이 432
앨머로드 제이
미국, 97355 오레곤, 레바논, 피.오. 박스 2692
스태드니스키 마이클 디.
미국, 97520 오레곤, 애쉬랜드, 모라다 레인 2397
(74) 대리인
강명구

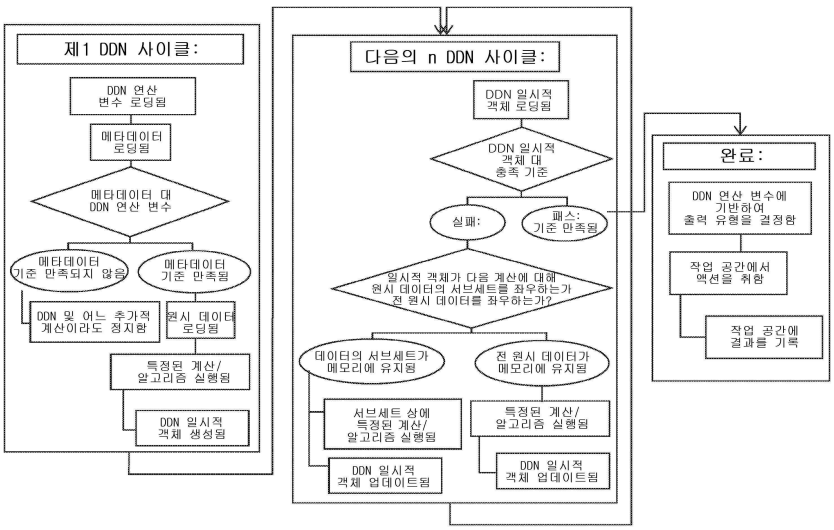
전체 청구항 수 : 총 79 항

(54) 발명의 명칭 **데이터 발견 노드**

(57) 요 약

외부 알고리즘을 인보크 및 동화하고 상기 알고리즘과 인-세션 및 실시간으로 상호작용하기 위한 프레임워크 및 인터페이스가 여기에서 설명된다. 일례의 실시형태는 또한 데이터 자체가 알고리즘 선택, 변수, 및 제시를 지시하여 분석 작업 흐름에서 반복 및 최적화를 가져올 수 있게 하는 데이터-구동식 분석을 위해 레버리징될 수 있는 재현가능한, 업데이트가능한 노드를 포함한다. 예시적 실시형태로, 발견 또는 진단 프로세스 전체는 특정 데이터 세트 상에 실행될 수 있고, 그로써 동일한 발견 또는 진단 프로세스, 표현형 식별, 및 시각화가 장래 실험 상에서 반복되거나, 발행되거나, 유효성 검증되거나, 또는 다른 조사자와 공유될 수 있도록 발견 또는 진단 프로세스를 특정 데이터 세트로부터 분리할 수 있다.

대 표 도 - 도6



(52) CPC특허분류
G06F 9/44526 (2013.01)
G06N 5/022 (2013.01)
H04L 63/102 (2013.01)
H04L 67/10 (2013.01)
H04L 67/141 (2013.01)
H04L 67/143 (2013.01)
H04L 67/20 (2013.01)

명 세 서

청구범위

청구항 1

어느 특정 데이터 세트에라도 독립적인 모델에 따라 과학적 데이터를 프로세싱하기 위한 컴퓨터 프로그램 제품으로서,

비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하는 데이터 발견 노드 데이터 구조로서, 상기 데이터 발견 노드 데이터 구조는 (1) 반복적 과학적 데이터 분석을 받게 될 과학적 데이터의 규격, (2) 상기 반복적 과학적 데이터 분석에 대한 출력 포맷의 규격, 및 (3) 상기 반복적 과학적 데이터 분석을 제어하기 위한 복수의 연산 변수의 규격을 포함하고, 특정된 상기 연산 변수는 (i) 상기 반복적 과학적 데이터 분석의 일부분으로서 특정된 상기 과학적 데이터 상에 수행될 알고리즘의 규격, (ii) 메타데이터의 규격으로서, 특정된 상기 메타데이터는 특정된 상기 알고리즘이 특정된 상기 과학적 데이터에 적용될 조건을 정의하도록 구성된, 상기 메타데이터의 규격, 및 (iii) 충족 변수의 규격으로서, 특정된 상기 충족 변수는 상기 반복적 과학적 데이터 분석의 일부분으로서 얼마나 많은 반복이 수행되는지 제어하도록 구성된, 상기 충족 변수의 규격을 포함하는, 상기 데이터 발견 노드 데이터 구조; 및

비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하는 복수의 프로세서-실행가능한 명령어를 포함하되, 상기 명령어는, 컴퓨터의 프로세서에 의한 실행시, 상기 컴퓨터로 하여금 상기 데이터 발견 데이터 구조를 판독 및 인보크하여 특정된 상기 연산 변수에 따라 특정된 상기 과학적 데이터에 대응하는 특정 데이터 세트 상에 상기 반복적 과학적 데이터 분석을 수행하고 특정된 상기 출력 포맷으로 결과를 발생시키게 하도록 구성되는, 컴퓨터 프로그램 제품.

청구항 2

제1항에 있어서, 상기 명령어는, 상기 프로세서에 의한 실행시, 상기 컴퓨터로 하여금, 상기 판독 및 인보크 연산의 일부분으로서,

특정된 상기 연산 변수를 메모리에 로딩하고;

로딩된 상기 메타데이터를 상기 특정 데이터 세트에 대비하여 시험하고;

상기 메타데이터 시험 연산에 기반하여 상기 특정 데이터 세트 상에 수행될 특정된 알고리즘을 결정하고;

결정된 상기 알고리즘을 상기 특정 데이터 세트에 적용하여 데이터 분석 결과를 산출하고;

상기 반복적 과학적 데이터 분석에 대한 상태 및 상기 결과를 저장하는 메모리에서의 일시적 데이터 객체를 생성하고;

로딩된 상기 충족 변수에 기반하여 상기 반복적 과학적 데이터 분석의 다른 반복이 필요한지 결정하고;

상기 반복적 과학적 데이터 분석의 다른 반복이 필요하다는 결정에 응답하여, (1) 로딩된 상기 충족 변수가 충족되었다는 결정이 이루어질 때까지 상기 메타데이터 시험 연산, 특정된 상기 알고리즘 결정, 상기 알고리즘 적용 연산, 및 상기 다른 반복 결정 연산을 반복하고, 그리고 (2) 반복된 상기 알고리즘 적용 연산에 기반하여 상기 일시적 데이터 객체를 업데이트하고; 그리고

상기 과학적 데이터 분석의 다른 반복이 필요하지 않다는 결정에 응답하여, 상기 반복적 과학적 데이터 분석의 결과를 특정된 상기 출력 포맷에 따라 작업 공간에 기록하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 3

제1항 또는 제2항에 있어서, 상기 비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하는 복수의 상기 데이터 발견 노드 데이터 구조를 더 포함하되, 복수의 상기 데이터 발견 노드 데이터 구조는 서로에 비해 다른 규격을 포함하고, 그리고 상기 명령어는, 상기 프로세서에 의한 실행시, 상기 컴퓨터로 하여금

사용자 입력에 응답하여 복수의 상기 데이터 발견 노드 데이터 구조로부터 데이터 발견 노드 데이터 구조를 선

택하고; 그리고

선택된 상기 데이터 발견 노드 데이터 구조 상에 상기 판독 및 인보크 연산을 수행하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 4

제3항에 있어서, 상기 명령어는, 상기 프로세서에 의한 실행시, 상기 컴퓨터로 하여금

사용자 입력에 응답하여 선택된 상기 데이터 발견 노드 데이터 구조에 대해 특정된 상기 연산 변수 중 적어도 하나를 정의하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 특정된 상기 알고리즘은 외부 알고리즘을 포함하고, 그리고 상기 명령어는, 상기 프로세서에 의한 실행시, 상기 컴퓨터로 하여금

외부 알고리즘 플러그-인 인터페이스 프레임워크를 통하여 상기 외부 알고리즘을 인보크하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 특정된 상기 알고리즘은 원격 컴퓨터에 의해 실행되고, 그리고 상기 명령어는, 상기 프로세서에 의한 실행시, 상기 컴퓨터로 하여금

원격 컴퓨터 플러그-인 인터페이스 프레임워크를 통하여 특정된 상기 알고리즘을 인보크하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 과학적 데이터는 실험으로부터의 세포 데이터를 포함하는, 컴퓨터 프로그램 제품.

청구항 8

제7항에 있어서, 상기 세포 데이터는 흐름 세포 측정기에 의해 획득된 단일 세포 데이터를 포함하는, 컴퓨터 프로그램 제품.

청구항 9

제7항 또는 제8항에 있어서, 특정된 상기 알고리즘은 집단 식별 알고리즘을 포함하는, 컴퓨터 프로그램 제품.

청구항 10

제7항 내지 제9항 중 어느 한 항에 있어서, 특정된 상기 알고리즘은 기하학적 게이팅 알고리즘을 포함하는, 컴퓨터 프로그램 제품.

청구항 11

과학적 데이터를 분석하기 위한 방법으로서,

데이터 발견 노드 데이터 구조를 데이터 파일에 적용하는 단계를 포함하고, 상기 데이터 파일은 획득 기기에 의해 수집된 과학적 데이터를 포함하고, 상기 데이터 파일은 상기 데이터 파일과 연관된 메타데이터를 갖되, 상기 적용하는 단계는

상기 데이터 파일과 연관된 상기 메타데이터 및 상기 데이터 발견 노드와 연관된 복수의 연산 변수를 메모리에 로딩하는 단계;

상기 메타데이터가 상기 복수의 연산 변수 중 하나에 의해 특정된 메타데이터 규칙 기준을 만족하는지 결정하는 단계; 및

상기 메타데이터가 상기 메타데이터 규칙 기준을 만족한다는 결정에 응답하여,

상기 데이터 파일과 연관된 상기 과학적 데이터를 메모리에 로딩하는 단계;

상기 데이터 파일과 연관된 상기 과학적 데이터 상에 제1 분석 알고리즘을 실행하는 단계로서, 상기 복수의 연산 변수 중 하나는 상기 제1 분석 알고리즘을 특정하는 상기 실행하는 단계;

충족 변수를 정의하는 일시적 데이터 객체를 생성하는 단계;

상기 일시적 데이터 객체의 충족 변수가 상기 복수의 연산 변수 중 하나에 의해 특정된 충족 임계치를 충족하는지 결정하는 단계; 및

상기 일시적 데이터 객체의 충족 변수가 상기 충족 임계치를 충족하지 않는다는 결정에 응답하여, (1) 상기 데이터 파일과 연관된 상기 과학적 데이터의 전 세트 또는 서브세트 상에 상기 제1 분석 알고리즘 또는 제2 분석 알고리즘 중 어느 하나를 실행하되, 상기 복수의 연산 변수 중 하나는 원시 데이터의 상기 전 세트 또는 상기 서브세트에 상기 제1 분석 알고리즘 또는 상기 제2 분석 알고리즘을 적용할지 정의하고, 그리고 (2) 상기 제1 분석 알고리즘 또는 상기 제2 분석 알고리즘의 실행에 기반하여 상기 일시적 데이터 객체를 업데이트하는 단계; 및

업데이트된 상기 일시적 데이터 객체의 충족 변수가 상기 충족 임계치를 충족할 때까지 (1) 상기 일시적 데이터 객체의 충족 변수가 상기 충족 임계치를 충족하는지 결정하는 단계, (2) 상기 제1 분석 알고리즘 또는 상기 제2 분석 알고리즘 중 어느 하나를 실행하는 단계, 및 (3) 상기 일시적 데이터 객체를 업데이트하는 단계를 반복적으로 수행하는 단계를 포함하되,

상기 방법 단계는 프로세서에 의해 수행되는, 과학적 데이터를 분석하기 위한 방법.

청구항 12

제11항에 있어서, 상기 데이터 발견 노드는 데이터베이스에 저장된 미리-정의된 데이터 구조인, 과학적 데이터를 분석하기 위한 방법.

청구항 13

제11항에 있어서, 상기 데이터 발견 노드는 사용자에 의해 생성된 새로 정의된 데이터 구조인, 과학적 데이터를 분석하기 위한 방법.

청구항 14

제13항에 있어서, 새로 정의된 상기 데이터 발견 노드를 장래 사용을 위해 데이터베이스에 저장하는 단계를 더 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 15

제14항에 있어서,

상기 프로세서가 상기 데이터베이스에서 유사한 데이터 발견 노드를 찾도록 검색하는 단계; 및

상기 유사한 데이터 발견 노드의 파라미터를 분석하고 상기 유사한 데이터 발견 노드의 상기 파라미터를 상기 데이터 발견 노드의 상기 파라미터와 비교하는 단계를 더 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 16

제11항 내지 제15항 중 어느 한 항에 있어서, 상기 연산 변수는 복수의 획득 기기 파라미터, 특징 변수, 반복 변수, 및 범위 변수를 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 17

제11항 내지 제16항 중 어느 한 항에 있어서, 상기 메타데이터가 메타데이터 규칙 기준을 만족하는지 결정하는 단계는 상기 프로세서가 느슨한 모드, 보통 모드, 및 엄격한 모드로 이루어진 그룹으로부터 선택된 모드에 따라 상기 메타데이터 규칙 기준에 대비하여 상기 메타데이터를 시험하는 단계를 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 18

컴퓨터 프로그램 제품으로서,

비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하는 복수의 프로세서-실행가능한 명령어를 포함하되, 상기 명령어는 컴퓨터로 하여금

노드 데이터 구조를 데이터 파일에 적용하게 함으로써 과학적 데이터를 분석하도록 프로세서에 의한 실행을 위해 구성되고, 상기 데이터 파일은 획득 기기에 의해 수집된 과학적 데이터를 포함하고, 상기 데이터 파일은 상기 데이터 파일과 연관된 메타데이터를 갖되, 상기 적용 연산은

상기 데이터 파일과 연관된 상기 메타데이터 및 상기 데이터 발견 노드와 연관된 복수의 연산 변수를 메모리에 로딩하고;

상기 메타데이터가 상기 복수의 연산 변수 중 하나에 의해 특정된 메타데이터 규칙 기준을 만족하는지 결정하고; 그리고

상기 메타데이터가 상기 메타데이터 규칙 기준을 만족한다는 결정에 응답하여,

상기 데이터 파일과 연관된 상기 과학적 데이터를 메모리에 로딩하고;

상기 데이터 파일과 연관된 상기 과학적 데이터 상에 제1 분석 알고리즘을 실행하되, 상기 복수의 연산 변수 중 하나는 상기 제1 분석 알고리즘을 특정하고;

충족 변수를 정의하는 일시적 데이터 객체를 생성하고;

상기 일시적 데이터 객체의 충족 변수가 상기 복수의 연산 변수 중 하나에 의해 특정된 충족 임계치를 충족하는지 결정하고; 그리고

상기 일시적 데이터 객체의 충족 변수가 상기 충족 임계치를 충족하지 않는다는 결정에 응답하여, (1) 상기 데이터 파일과 연관된 상기 과학적 데이터의 전 세트 또는 서브세트 상에 상기 제1 분석 알고리즘 또는 제2 분석 알고리즘 중 어느 하나를 실행하되, 상기 복수의 연산 변수 중 하나는 원시 데이터의 상기 전 세트 또는 상기 서브세트에 상기 제1 분석 알고리즘 또는 상기 제2 분석 알고리즘을 적용할지 정의하고, 그리고 (2) 상기 제1 분석 알고리즘 또는 상기 제2 분석 알고리즘의 실행에 기반하여 상기 일시적 데이터 객체를 업데이트하고; 그리고

업데이트된 상기 일시적 데이터 객체의 충족 변수가 상기 충족 임계치를 충족할 때까지 (1) 상기 일시적 데이터 객체의 충족 변수가 상기 충족 임계치를 충족하는지에 관한 결정 연산, (2) 상기 제1 분석 알고리즘 또는 상기 제2 분석 알고리즘 실행 연산, 및 (3) 상기 업데이트 연산을 반복적으로 수행하도록 구성되는, 컴퓨터 프로그램 제품.

청구항 19

과학적 데이터를 분석하기 위한 방법으로서,

복수의 연산 변수의 규격을 수신하는 단계로서, 상기 규격은 (1) 충족 기준의 규격, (2) 제1 분석 알고리즘의 규격, (3) 제2 분석 알고리즘의 규격, 및 (4) 상기 제1 분석 알고리즘 및 상기 제2 분석 알고리즘이 상기 과학적 데이터에 적용되어야 하는 조건의 규격을 포함하는 상기 수신하는 단계;

상기 제1 분석 알고리즘을 특정하는 상기 연산 변수 및 상기 제1 분석 알고리즘이 상기 과학적 데이터에 적용되어야 하는 상기 조건을 특정하는 상기 연산 변수에 기반하여 상기 과학적 데이터의 적어도 일부 상에 상기 제1 분석 알고리즘을 실행하는 단계; 및

상기 충족 기준이 만족될 때까지 상기 연산 변수 및 상기 실행하는 단계의 결과에 기반하여 상기 과학적 데이터의 적어도 일부 상에 상기 제1 분석 알고리즘 또는 제2 분석 알고리즘을 반복적으로 실행하는 단계를 포함하되,

상기 방법 단계는 프로세서에 의해 수행되는, 과학적 데이터를 분석하기 위한 방법.

청구항 20

제19항에 있어서, 상기 복수의 연산 변수 중 하나는 상기 과학적 데이터를 분석할지 상기 과학적 데이터의 서브세트를 분석할지 특정하는, 과학적 데이터를 분석하기 위한 방법.

청구항 21

제19항 또는 제20항에 있어서, 상기 프로세서가 충족 값을 포함하는 일시적 객체를 생성하는 단계로서, 상기 프로세서는 상기 프로세서가 데이터 상에 분석 알고리즘을 실행한 후에 상기 충족 값을 업데이트하는 상기 생성하는 단계; 및

상기 프로세서가 상기 충족 기준이 만족되는지 결정하기 위해 상기 충족 값의 값을 상기 충족 기준에 비교하는 단계를 더 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 22

제19항 내지 제21항 중 어느 한 항에 있어서, 상기 프로세서는 상기 프로세서가 상기 과학적 데이터 상에 상기 제1 분석 알고리즘을 실행함으로써 발생된 결과에 기반하여 상기 제1 분석 알고리즘을 실행할지 상기 제2 분석 알고리즘을 실행할지 결정하는, 과학적 데이터를 분석하기 위한 방법.

청구항 23

제19항 내지 제22항 중 어느 한 항에 있어서, 상기 프로세서는 상기 프로세서가 상기 과학적 데이터 상에 상기 제1 분석 알고리즘을 실행함으로써 발생된 결과에 기반하여 상기 과학적 데이터를 분석할지 상기 과학적 데이터의 서브세트를 분석할지 결정하는, 과학적 데이터를 분석하기 위한 방법.

청구항 24

제19항 내지 제23항 중 어느 한 항에 있어서, 특정된 상기 분석 알고리즘 중 적어도 하나는 외부 분석 알고리즘을 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 25

방법으로서,

메모리에서의 데이터 발견 노드 데이터 구조에 액세스하는 단계로서, 상기 데이터 발견 노드 데이터 구조는 복수의 데이터 세트 중 어느 하나 상에 수행될 복수의 프로세싱 연산을 정의하도록 구성되되, 상기 데이터 발견 노드 데이터 구조는 어느 특정 데이터 세트에라도 독립적인 프로세싱을 위한 모델로서 구성되는 상기 액세스하는 단계;

데이터 세트를 선택하는 단계; 및

프로세서에 의해, 선택된 상기 데이터 세트에 대한 프로세싱 결과를 발생시키기 위해 액세스된 상기 데이터 발견 노드 데이터 구조를 선택된 상기 데이터 세트 내 데이터에 적용하는 단계를 포함하는 방법.

청구항 26

제25항에 있어서, 상기 데이터 발견 노드 데이터 구조는 (1) 충족 기준의 규격, (2) 제1 분석 알고리즘의 규격, (3) 제2 분석 알고리즘의 규격, 및 (4) 상기 제1 분석 알고리즘 및 상기 제2 분석 알고리즘이 데이터 세트에 적용되어야 하는 조건의 규격을 포함하고; 그리고

상기 적용하는 단계는 상기 프로세서가 (1) 특정된 상기 분석 알고리즘 중 어느 것을 실행할지 특정된 상기 조건에 기반하여 결정하는 단계, (2) 선택된 상기 데이터 세트 내 데이터에 대해 결정된 상기 분석 알고리즘을 실행하는 단계, (3) 결정된 상기 분석 알고리즘의 실행이 특정된 상기 충족 기준의 충족을 초래하였는지 결정하는 단계, 및 (4) 특정된 상기 충족 기준이 충족되었다는 결정이 이루어질 때까지 (i) 특정된 상기 분석 알고리즘 중 어느 것을 실행할지 특정된 상기 조건에 기반하여 결정하는 단계, (ii) 선택된 상기 데이터 세트 내 데이터에 대해 결정된 상기 분석 알고리즘을 실행하는 단계, 및 (iii) 결정된 상기 분석 알고리즘의 실행이 특정된 상기 충족 기준의 충족을 초래하였는지 결정하는 단계를 반복하는 단계를 포함하는 방법.

청구항 27

컴퓨터 프로그램 제품으로서,

비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하는 복수의 프로세서-실행가능한 명령어를 포함하되, 상기

명령어는 컴퓨터로 하여금

복수의 연산 변수의 규격을 수신하되, 상기 규격은 (1) 충족 기준의 규격, (2) 제1 분석 알고리즘의 규격, (3) 제2 분석 알고리즘의 규격, 및 (4) 상기 제1 분석 알고리즘 및 상기 제2 분석 알고리즘이 과학적 데이터에 적용되어야 하는 조건의 규격을 포함하고;

상기 제1 분석 알고리즘을 특정하는 상기 연산 변수 및 상기 제1 분석 알고리즘이 상기 과학적 데이터에 적용되어야 하는 상기 조건을 특정하는 상기 연산 변수에 기반하여 상기 과학적 데이터의 적어도 일부 상에 상기 제1 분석 알고리즘을 실행하고; 그리고

상기 충족 기준이 만족될 때까지 상기 연산 변수 및 상기 실행 단계의 결과에 기반하여 상기 과학적 데이터의 적어도 일부 상에 상기 제1 분석 알고리즘 또는 제2 분석 알고리즘을 반복적으로 실행하게 함으로써 상기 과학적 데이터를 분석하도록 프로세서에 의한 실행을 위해 구성되는, 컴퓨터 프로그램 제품.

청구항 28

컴퓨터 프로그램 제품으로서,

비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하는 복수의 프로세서-실행가능한 명령어를 포함하되, 상기 명령어는 컴퓨터로 하여금

메모리에서의 데이터 발견 노드 데이터 구조에 액세스하고, 상기 데이터 발견 노드 데이터 구조는 복수의 데이터 세트 중 어느 하나 상에 수행될 복수의 프로세싱 연산을 정의하도록 구성되되, 상기 데이터 발견 노드 데이터 구조는 어느 특정 데이터 세트에라도 독립적인 프로세싱을 위한 모델로서 구성되고;

데이터 세트를 선택하고; 그리고

프로세서에 의해, 선택된 상기 데이터 세트에 대한 프로세싱 결과를 발생시키기 위해 액세스된 상기 데이터 발견 노드 데이터 구조를 선택된 상기 데이터 세트 내 데이터에 적용하게 함으로써 과학적 데이터를 분석하도록 상기 프로세서에 의한 실행을 위해 구성되는, 컴퓨터 프로그램 제품.

청구항 29

제28항에 있어서, 상기 데이터 발견 노드 객체는 라이브러리로서 저장된 데이터 발견 객체의 데이터베이스에서의 복수의 데이터 발견 노드 객체 중 하나를 포함하되, 각각의 데이터 발견 객체는 선택된 상기 데이터에 기반하여 분석 알고리즘을 특정하도록 구성된 메타데이터와 연관되는, 컴퓨터 프로그램 제품.

청구항 30

제28항 또는 제29항에 있어서, 상기 데이터 세트는 흐름 세포 측정기에 의해 획득된 단일 세포 데이터를 포함하는, 컴퓨터 프로그램 제품.

청구항 31

컴퓨터에 의해 실행되는 데이터 분석 소프트웨어와 상기 데이터 분석 소프트웨어의 외부에 있는 알고리즘 간 플러그-인 인터페이스 프레임워크를 제공하기 위한 컴퓨터 프로그램 제품으로서,

비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하고 상기 플러그-인 인터페이스 프레임워크를 정의하는 복수의 프로세서-실행가능한 명령어를 포함하되, 상기 명령어는, 상기 컴퓨터의 프로세서에 의한 실행시, 상기 컴퓨터로 하여금

상기 데이터 분석 소프트웨어의 실행의 세션 동안 샘플 데이터의 분석과 관련된 작업 공간을 열고;

상기 외부 알고리즘으로의 인터페이스를 위한 랩퍼를 생성하고;

상기 외부 알고리즘의 인보크에 대한 요청을 구성하고, 상기 요청은 상기 외부 알고리즘의 식별 및 상기 외부 알고리즘에 대한 데이터 입력의 규격을 포함하고;

상기 요청에 따라 상기 외부 알고리즘을 인보크하기 위해 샘플 데이터, 상기 샘플 데이터에 대한 메타데이터, 및 상기 외부 알고리즘이 결과를 저장해야 하는 출력 목적지를 상기 외부 알고리즘에 제공하고;

상기 출력 목적지에서 상기 외부 알고리즘으로부터 결과를 수신하고;

수신된 상기 결과를 상기 작업 공간에 저장하고; 그리고

상기 데이터 분석 소프트웨어에 대한 상기 작업 공간을 복원하여 상기 외부 알고리즘으로부터의 상기 결과를 상기 데이터 분석 소프트웨어 작업 공간에 통합하게 하도록 구성되는, 컴퓨터 프로그램 제품.

청구항 32

제31항에 있어서, 상기 샘플 데이터는 실험으로부터의 세포 데이터를 포함하는, 컴퓨터 프로그램 제품.

청구항 33

제32항에 있어서, 상기 세포 데이터는 흐름 세포 측정기에 의해 획득된 단일 세포 데이터를 포함하는, 컴퓨터 프로그램 제품.

청구항 34

제32항 또는 제33항에 있어서, 상기 플러그-인 인터페이스 프레임워크는 (1) 상기 샘플 데이터의 입력 파일, (2) 상기 출력 목적지, 및 (3) 상기 샘플 데이터의 데이터 기술을 포함하는, 컴퓨터 프로그램 제품.

청구항 35

제34항에 있어서, 상기 데이터 기술은 (1) 상기 샘플 데이터에 대한 메타데이터, (2) 상기 샘플 데이터에 도달하는데 사용된 하나 이상의 이전 계산의 식별, 및 (3) 상기 샘플 데이터를 프로세싱 및 시각화하는데 사용되는 하나 이상의 변환의 식별을 포함하는, 컴퓨터 프로그램 제품.

청구항 36

제32항 내지 제35항 중 어느 한 항에 있어서, 상기 명령어는, 상기 프로세서에 의한 실행시, 상기 컴퓨터로 하여금

상기 외부 알고리즘에 의해 발생된 아티팩트를 관리하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 37

제32항 내지 제36항 중 어느 한 항에 있어서, 상기 명령어는 제1 언어로 작성되고, 상기 제1 언어는 제2 언어의 프로그램에 대한 호출을 지원하도록 구성되고, 그리고 상기 외부 알고리즘은 상기 제2 언어로 작성되는, 컴퓨터 프로그램 제품.

청구항 38

제36항에 있어서, 상기 제1 언어는 자바인, 컴퓨터 프로그램 제품.

청구항 39

제38항에 있어서, 상기 제2 언어는 C인, 컴퓨터 프로그램 제품.

청구항 40

제38항에 있어서, 상기 제2 언어는 R인, 컴퓨터 프로그램 제품.

청구항 41

제38항에 있어서, 상기 제2 언어는 웹-기반 계산 엔진 언어인, 컴퓨터 프로그램 제품.

청구항 42

제32항 내지 제41항 중 어느 한 항에 있어서, 상기 랩퍼는 외부 집단 노드를 포함하는, 컴퓨터 프로그램 제품.

청구항 43

제42항에 있어서, 상기 명령어는, 상기 프로세서에 의한 실행시, 상기 컴퓨터로 하여금

상기 샘플 데이터에 대한 유도된 파라미터를 생성하게 하도록 더 구성되고, 상기 유도된 파라미터는 상기 샘플

데이터 내 아집단을 생성하도록 상기 데이터 분석 소프트웨어에 의한 사용을 위한 것인, 컴퓨터 프로그램 제품.

청구항 44

제32항 내지 제43항 중 어느 한 항에 있어서, 상기 데이터 입력 규격은 (1) 상기 샘플 데이터에 대한 파일 경로, (2) 상기 샘플 데이터에 대한 파라미터의 수, (3) 상기 샘플 데이터에 대한 클러스터의 수, (4) 상기 샘플 데이터에 대한 차원 축소를 위한 변수, (5) 색 선택, (6) 시각화 유형, 및 (7) 이미지 유형으로 이루어진 그룹의 적어도 하나의 일원을 포함하는, 컴퓨터 프로그램 제품.

청구항 45

제32항 내지 제44항 중 어느 한 항에 있어서, 상기 외부 알고리즘은 집단 식별 알고리즘을 포함하는, 컴퓨터 프로그램 제품.

청구항 46

제32항 내지 제45항 중 어느 한 항에 있어서, 특정된 상기 알고리즘은 기하학적 게이팅 알고리즘을 포함하는, 컴퓨터 프로그램 제품.

청구항 47

제32항 내지 제46항 중 어느 한 항에 있어서, 특정된 상기 알고리즘은 보고서 발생 알고리즘을 포함하는, 컴퓨터 프로그램 제품.

청구항 48

제32항 내지 제47항 중 어느 한 항에 있어서, 특정된 상기 알고리즘은 분석 시각화 알고리즘을 포함하는, 컴퓨터 프로그램 제품.

청구항 49

제32항 내지 제49항 중 어느 한 항에 있어서, 특정된 상기 알고리즘은 높은 처리율 유전자형 분석 알고리즘 및 높은 처리율 표현형 데이터 분석 알고리즘으로 이루어진 그룹의 일원을 포함하는, 컴퓨터 프로그램 제품.

청구항 50

제32항 내지 제49항 중 어느 한 항에 있어서, 상기 명령어는 상기 데이터 분석 소프트웨어의 일부분인, 컴퓨터 프로그램 제품.

청구항 51

제32항 내지 제50항 중 어느 한 항에 있어서, 상기 명령어는 상기 데이터 분석 소프트웨어의 사용자에 의한 개입 없이 상기 컴퓨터의 상기 프로세서에 의한 실행을 위해 구성되는, 컴퓨터 프로그램 제품.

청구항 52

과학적 데이터를 분석하기 위한 방법으로서,

외부 집단 노드를 생성하는 단계로서, 상기 외부 집단 노드는 (1) 획득 기기에 의해 수집된 데이터를 표현하는 데이터 세트, (2) 출력 폴더 목적지, 및 (3) 상기 데이터 세트의 마크업 언어 기술을 포함하고, 상기 마크업 언어 기술은 (i) 상기 데이터 세트에 대한 메타-정보, 및 (ii) 상기 데이터 세트에 도달하는데 사용된 이전 계산의 규격을 포함하는, 상기 생성하는 단계;

상기 외부 집단 노드에 기반하여 외부 알고리즘을 인보크하는 단계로서, 상기 외부 알고리즘은 상기 데이터 세트 상에 프로세싱을 수행하도록 구성되고, 그리고 상기 출력 폴더 목적지는 상기 외부 알고리즘이 당해 프로세싱으로부터의 결과를 저장하여야 하는 곳을 정의하는, 상기 인보크하는 단계;

상기 데이터 세트 상에 상기 외부 알고리즘에 의해 수행된 상기 프로세싱의 결과로서 상기 외부 알고리즘으로부터의 결과를 수신하는 단계; 및

상기 외부 알고리즘으로부터의 상기 결과를 분석 작업 공간 파일에 통합하는 단계를 포함하되,

상기 방법 단계는 프로세서에 의해 수행되는, 과학적 데이터를 분석하기 위한 방법.

청구항 53

제52항에 있어서, 상기 외부 알고리즘은 세포 집단 식별 알고리즘을 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 54

제52항 또는 제53항에 있어서, 상기 외부 알고리즘은 높은-처리율 유전자형 데이터 분석 알고리즘을 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 55

제52항 내지 제54항 중 어느 한 항에 있어서, 상기 외부 알고리즘은 분석 결과의 시각화 및 보고서를 발생시키는, 과학적 데이터를 분석하기 위한 방법.

청구항 56

제52항 내지 제55항 중 어느 한 항에 있어서, 상기 생성하는 단계는 상기 프로세서가 상기 외부 알고리즘에 의해 사용될 연산 변수를 설정하라고 사용자에게 프롬프트하는 단계를 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 57

제52항 내지 제56항 중 어느 한 항에 있어서, 상기 외부 알고리즘 결과는 (1) CSV 파일, (2) CLR 파일, (3) 게이팅 ML 파일, 및 (4) FCS 파일로 이루어진 그룹의 일원을 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 58

제52항 내지 제57항 중 어느 한 항에 있어서, 상기 데이터 세트가 변경될 때 상기 프로세서가 상기 외부 알고리즘을 다시 인보크하는 단계를 더 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 59

과학적 데이터를 분석하기 위한 방법으로서,

(1) 획득 기기에 의해 수집된 과학적 데이터와 연관된 메타데이터 및 (2) 원격 컴퓨터 시스템에 접속하기 위한 인터페이스와 연관된 메타데이터를 판독하는 단계;

상기 과학적 데이터와 연관된 상기 메타데이터 및 상기 인터페이스와 연관된 상기 메타데이터에 기반하여 원시 데이터 분석의 방법을 지시하는 단계;

상기 인터페이스 및 작업 공간의 상태를 저장하는 단계;

상기 원격 컴퓨터 시스템에 대한 네트워크 주소를 참조로 하고 상기 원격 컴퓨터 시스템과의 접속을 발생시킴으로써 상기 인터페이스를 통해 상기 원격 컴퓨터 시스템과 통신하는 단계;

상기 작업 공간과 연관된 작업 공간 데이터를 상기 원격 컴퓨터 시스템에 송신하는 단계;

상기 원격 컴퓨터 시스템으로부터 데이터를 수신하는 단계; 및

상기 원격 컴퓨터 시스템과의 상기 접속을 종료하는 단계를 포함하되,

상기 방법 단계는 프로세서에 의해 수행되는, 과학적 데이터를 분석하기 위한 방법.

청구항 60

제59항에 있어서, 상기 원격 컴퓨터 시스템으로부터 수신된 상기 데이터는 (1) 오픈-소스 데이터베이스로부터의 실험 데이터, (2) 외부 데이터베이스로부터의 주석 달린 실험 데이터, 및 (3) 임상 시험을 위한 추적 식별 번호로 이루어진 그룹의 일원을 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 61

제59항 또는 제60항에 있어서, 상기 종료하는 단계는 상기 프로세서가 데이터베이스 접속 또는 다른 컴퓨터 자원을 풀어주는 단계를 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 62

제59항 내지 제61항 중 어느 한 항에 있어서, 상기 원격 컴퓨터 시스템으로부터 권한 부여를 받기 위해 상기 원격 컴퓨터 시스템과 크리덴셜을 교환하는 단계를 더 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 63

제59항 내지 제62항 중 어느 한 항에 있어서, 상기 프로세서가 상기 원격 컴퓨터 시스템에 의해 수행된 프리-프로세싱 단계 및 포스트-프로세싱 단계로 이루어진 그룹의 일원을 반영하기 위해 상기 작업 공간과 연관된 상기 메타데이터를 증강하는 단계를 더 포함하는, 과학적 데이터를 분석하기 위한 방법.

청구항 64

제1 컴퓨터에 의해 실행되는 데이터 분석 소프트웨어가 제2 컴퓨터에 의해 실행되는 연산 함수를 인보크할 수 있게 하도록 상기 제1 컴퓨터와 상기 제2 컴퓨터 간 플러그-인 인터페이스 프레임워크를 제공하기 위한 컴퓨터 프로그램 제품으로서,

비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하고 상기 플러그-인 인터페이스 프레임워크를 정의하는 복수의 프로세서-실행가능한 명령어를 포함하되, 상기 명령어는, 상기 제1 컴퓨터의 프로세서에 의한 실행시, 상기 제1 컴퓨터로 하여금

상기 데이터 분석 소프트웨어의 실행의 세션 동안 샘플 데이터의 분석과 관련된 작업 공간을 열고;

상기 작업 공간의 데이터 기술을 판독하되, 상기 작업 공간 데이터 기술은 상기 샘플 데이터에 대한 메타데이터를 포함하고;

상기 플러그-인 인터페이스 프레임워크의 데이터 기술을 판독하되, 상기 인터페이스 데이터 기술은 상기 제2 컴퓨터에 대한 주소를 식별시키는 메타데이터를 포함하고;

상기 주소에 기반하여 상기 제2 컴퓨터로의 접속을 확립하고;

상기 샘플 데이터에 대해 상기 제2 컴퓨터의 연산 함수를 인보크하기 위해 상기 접속을 통해 상기 제2 컴퓨터와 통신하고;

상기 인보크에 응답하여 상기 접속을 통해 상기 제2 컴퓨터로부터 데이터를 수신하고;

수신된 상기 데이터에 기반하여 상기 작업 공간을 수정하고;

수정된 상기 작업 공간에 기반하여 상기 작업 공간 메타데이터를 수정하고;

수정된 상기 작업 공간을 상기 데이터 분석 소프트웨어 세션 내에서 메모리에 저장하고;

수정된 상기 작업 공간 메타데이터를 메모리에 저장하고; 그리고

상기 세션을 종료하게 하도록 구성되는, 컴퓨터 프로그램 제품.

청구항 65

제64항에 있어서, 상기 샘플 데이터는 실험으로부터의 세포 데이터를 포함하는, 컴퓨터 프로그램 제품.

청구항 66

제65항에 있어서, 상기 세포 데이터는 흐름 세포 측정기에 의해 획득된 단일 세포 데이터를 포함하는, 컴퓨터 프로그램 제품.

청구항 67

제65항 또는 제66항에 있어서, 상기 명령어는, 상기 제1 컴퓨터의 상기 프로세서에 의한 실행시, 상기 제1 컴퓨

터로 하여금

상기 세포 데이터에 대해 게이트의 부가를 통하여 수신된 상기 데이터에 기반하여 상기 작업 공간을 수정하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 68

제65항 내지 제67항 중 어느 한 항에 있어서, 상기 명령어는, 상기 제1 컴퓨터의 상기 프로세서에 의한 실행시, 상기 제1 컴퓨터로 하여금

상기 세포 데이터에 대한 통계의 부가를 통하여 수신된 상기 데이터에 기반하여 상기 작업 공간을 수정하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 69

제65항 내지 제68항 중 어느 한 항에 있어서, 상기 명령어는, 상기 제1 컴퓨터의 상기 프로세서에 의한 실행시, 상기 제1 컴퓨터로 하여금

상기 세포 데이터에 대한 샘플 이름의 부가를 통하여 수신된 상기 데이터에 기반하여 상기 작업 공간을 수정하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 70

제65항 내지 제69항 중 어느 한 항에 있어서, 상기 명령어는, 상기 제1 컴퓨터의 상기 프로세서에 의한 실행시, 상기 제1 컴퓨터로 하여금

수정된 상기 작업 공간이 정의된 스키마를 고수하게 되도록 수신된 상기 데이터에 기반하여 상기 작업 공간을 수정하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 71

제65항 내지 제70항 중 어느 한 항에 있어서, 상기 명령어는, 상기 제1 컴퓨터의 상기 프로세서에 의한 실행시, 상기 제1 컴퓨터로 하여금

수정된 상기 작업 공간에 기반하여 상기 데이터 분석 소프트웨어에 의해 계산을 인보크하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 72

제65항 내지 제71항 중 어느 한 항에 있어서, 상기 명령어는, 상기 제1 컴퓨터의 상기 프로세서에 의한 실행시, 상기 제1 컴퓨터로 하여금

수정된 상기 작업 공간 메타데이터가 상기 세션에 대해 상기 플러그-인 인터페이스 프레임워크에 대한 보유된 상태를 표현하는 데이터를 포함하게 되도록 수정된 상기 작업 공간에 기반하여 상기 작업 공간 메타데이터를 수정하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 73

제65항 내지 제72항 중 어느 한 항에 있어서, 상기 명령어는, 상기 제1 컴퓨터의 상기 프로세서에 의한 실행시, 상기 제1 컴퓨터로 하여금

컴퓨터 자원을 풀어주도록 상기 접속의 종료를 통해 상기 세션을 종료하게 하도록 더 구성되는, 컴퓨터 프로그램 제품.

청구항 74

제65항 내지 제73항 중 어느 한 항에 있어서, 상기 샘플 데이터는 복수의 파라미터에 대한 데이터 값을 포함하고, 그리고 상기 샘플 데이터 메타데이터는 상기 샘플 데이터에 대한 파라미터 이름을 포함하는, 컴퓨터 프로그램 제품.

청구항 75

제65항 내지 제74항 중 어느 한 항에 있어서, 상기 제2 컴퓨터 주소는 데이터베이스 또는 서버에 대한 URI를 포함하는, 컴퓨터 프로그램 제품.

청구항 76

제65항 내지 제75항 중 어느 한 항에 있어서, 상기 명령어는 상기 데이터 분석 소프트웨어의 일부분인, 컴퓨터 프로그램 제품.

청구항 77

제65항 내지 제76항 중 어느 한 항에 있어서, 상기 명령어는 상기 데이터 분석 소프트웨어의 사용자에 의한 개입 없이 상기 제1 컴퓨터의 상기 프로세서에 의한 실행을 위해 구성되는, 컴퓨터 프로그램 제품.

청구항 78

제65항 내지 제77항 중 어느 한 항에 있어서, 상기 작업 공간은 XML 데이터를 포함하는, 컴퓨터 프로그램 제품.

청구항 79

제65항 내지 제78항 중 어느 한 항에 있어서, 상기 작업 공간 데이터 기술은 XML 데이터를 포함하는, 컴퓨터 프로그램 제품.

발명의 설명

기 술 분 야

[0001] 관련 특허 출원에 대한 상호-참조 및 우선권 주장

[0002] 본 특허 출원은 2015년 5월 8일자로 출원된 미국 가특허출원 일련 번호 제62/158,903호 발명의 명칭 "Data Discovery Nodes"에 대한 우선권을 주장하며, 그 개시 전체는 참고로 본 명세서에 편입된다.

배 경 기 술

[0003] 기술 개량에 기인하여, 단일 세포 실험 기기는 이전 기기 세대보다 훨씬 더 많은 정보를 발생시킬 수 있다. 예컨대, 흐름 세포 측정기는, 각각의 세포에 대해 수많은 파라미터(예컨대, 10개 이상의 파라미터)로, 많은 수천의 개개의 세포를 표현하는 데이터를 발생시킬 수 있다. 결과적으로, 잠재적으로 식별될 수 있는 표현형의 수는 기하급수적으로 증가하였다. 환언하면, 단일 세포 검정에 의해 산출된 정보 콘텐츠는 실질적으로 본 출원시 이전에 증가하여 왔다. 부가적으로, 단일 세포 탐구는 세포당 많은 수천의 전사(RNA) 분자 및 DNA 수식의 질의를 포함하도록 확장되었다. 예컨대, 전체 전사체 분석은 한 번에 10,000개의 유전자를 조사할 것이다.

[0004] 더 많은 데이터를 발생시키는 것은 세포 표현형이 서로 상호작용하거나 질병에 영향을 미치는 방식 및 다른 질병-관련 단백질을 표현할 그들의 잠재력에 대한 더 많은 통찰을 제공하지만, 획득 기기에 의해 발생된 데이터의 가파른 양은 경이적일 수 있고 유수의 전문가조차 압도할 수 있다. 일반적으로, 생명 과학자는 그들의 전문성을 세포 기능 또는 세포 표현형 세트 또는 서브-세트에 초점 맞춘다. 예컨대, 면역학자는 질병 또는 면역 세포 기능을 이해할 한 줌의 세포 표현형에 그 또는 그녀의 실무를 초점 맞춘다. 한편, 시스템-생물학자는 세포 상호작용 및 유전자와 단백질을 함께 링크하는 경로에 대한 풍부한 지식을 가지고 있을 수 있다. 세포 상호작용, 식별, 및 기능성은 다양한 그리고 복잡한 범위의 속성을 포함하기 때문에 개개인이 모든 세포 집단(cell population)에서 전문가이기를 기대하는 것은 비현실적이다. 생명 과학자의 전문성은 일반적으로, 전부가 아니라, (보통은 모든 현재 알려져 있는 세포 표현형의 50%보다 더 적은) 일부의 세포 표현형으로 한정되기 때문에, 발견 및 진단 분석에서 지식 부조화가 생성되는데 전문가가 각각의 세포 표현형이 질병 또는 세포 상호작용에 어떻게 상관되는지를 상세하게 알지 못하기 때문이다. 이러한 지식 부조화의 결과로서, 전문가는 획득 기기에 의해 획득된 데이터에 대한 그의 연구를 그 전문가가 강력하게 알고 있는 세포 표현형에 초점 맞출 수 있다. 실험 및 연구를 표현형의 서브세트로 한정시, 분석가는 질병 또는 세포 기능에 매우 큰 영향력을 가지고 있을 수 있는 중요한 표현형을 무시하거나 놓칠 수 있다. 더욱, 알고 있는 표현형에 초점 맞춤으로써, 획득 기기에 의해 수집된 다량의 데이터는 휴면 또는 불용으로 있을 수 있다.

[0005] 세포 표현형의 서브세트에 기반하여 데이터의 서브세트를 분석하는 것은 실험 내에서 흥미로운 조사 결과를 가져올 수 있다. 그렇지만, 세포 반응은 다수 기능의 패턴을 표현하는 세포를 포함할 수 있고, 그리고, 세포 표현

형의 서브세트만을 분석함으로써, 과학자는 다른 세포 집단이 세포 반응 또는 질병에 어떻게 영향을 주는지 인식하지 못할 수 있다. 예컨대, 조사자는 특정 면역 반응에서 중요한 T-세포 서브세트를 찾는 실험을 수행하고 있을 수 있다. 이러한 예에서, T-세포 서브세트는 4개의 (마커라고도 알려져 있는) 파라미터 조합에 의해 정의될 수 있다. 물론, 실험의 시초에서, 조사자는 관심 있는 T-세포 서브세트를 식별하는데 필요한 마커의 수를 알고 있지 않다. 그리하여, 더 많은 세포에 대해 더 많은 마커를 조사함으로써, 조사자는 이환율 또는 치료 효능과 상관되는 세포 서브세트를 발견할 수 있고, 그리고, 더 많은 데이터 분석 기술로, 조사자는 그 자신의 지식 부조화를 극복하여 질병 또는 세포 기능에서 중요한 새로운 그리고 예상치 못한 서브세트를 찾아낼 수 있다. 그리하여, 대부분의 조사자 및 과학자에 의해 드러나는 지식 갭을 보상하는 기술 분야에서의 필요성이 존재한다.

[0006] 발명자는 관용적 기술 해법이 과학자의 지식 부족과 실제 세포 반응 간 갭을 충분히 가교하지 못한다고 생각한다. 예컨대, 관용적 기술은 가치 있는 분석 툴을 제공함으로써 조사자의 실험을 조력할 수 있기는 하지만, 그들 툴은 여전히 데이터-지식 부조화를 가교하기에는 충분하지 않다. 관용적 발견 해법에서, 분석가는 여전히 수동 클러스터링을 수행해야 하고 분석을 샘플 그룹에 적용해야 한다. 그렇지만, 세포 표현형을 조사하기 위한 9개의 마커, 메모리 상태를 조사하는 8개의 마커, 및 세포 시그널링을 조사하는 8개의 마커를 갖는 일례의 실험에서, 가능한 클러스터의 수는, 수동 분석에는 너무나 많은 클러스터인, 2^{25} 또는 33,554,432개의 클러스터이다. 환언하면, 잠재적 표현형의 수 및 가능한 2-차원 디스플레이는 수동 분석과 잘 스케일링하지 않는다. 물론, 표현형 공간을 더 관리가능한 수로 내려 한정하도록 소정 표현형 프루닝(pruning)이 일어날 수 있다. 예컨대, 생명 과학자는, 나이브(Naïve), 이펙터, 중앙 메모리, 및 이펙터 메모리 세포로 더 나뉘는, CD3+CD45+/-/HLA-DR-/CD16+, CD4+, 및 CD8+ 세포를 조사하기 위한 추가적 표현형 프루닝으로 단일 세포 및 라이브, 온전한 세포에 대한 프리-프로세스 게이팅을 수행할 수 있다. 그렇지만, 이러한 표현형-프루닝된 예에서조차도, 샘플당 16개의 파일의 수동 조작이 발견에 요구된다. 그리하여, 좁은 초점 너머로 발견-초점 맞춰진 연구에서 단일-세포 기술을 레버리징하려고 시도하는 과학자는 어려운, 비-결정론적, 그리고 비-재현가능한 경로와 직면한다. 그래서, 고도로-숙련된 전문가의 개입 없이 고차원 데이터를 분석하고 생물학적으로 관련 있는 데이터를 찾아낼 수 있는 데이터 분석 툴을 제공할 분야에서의 필요성이 존재한다.

발명의 내용

[0007] 위 문제점을 고려하여 본 발명이 개발되었다. 발명자는 어느 외부 알고리즘이라도 인보크 및 동화하고 상기 알고리즘과 인-세션 및 실시간으로 상호작용하기 위한 프레임워크 및 인터페이스를 개시한다. 발명자는 또한 재현가능한, 업데이트가능한 노드 및 이들 노드를, 데이터 자체가 알고리즘 선택, 변수, 및 제시를 지시하여 분석 작업 흐름에서 반복 및 최적화를 가져올 수 있게 하는 데이터-구동식 분석을 위해, 레버리징하는 것을 개시한다. 예시적 실시형태의 이들 2개의 태양을 통해, 발견 또는 진단 프로세스 전체는 특정 데이터 세트 상에 실행될 수 있고, 그로써 동일한 발견 또는 진단 프로세스, 표현형 식별, 및 시각화가 장래 실험 상에서 반복되거나, 발행되거나, 유효성 검증되거나, 또는 다른 조사자와 공유될 수 있도록 발견 또는 진단 프로세스를 특정 데이터 세트로부터 분리할 수 있다.

[0008] 본 발명의 추가적 특징 및 이점은 물론, 본 발명의 다양한 실시형태의 구조 및 동작 또한, 수반 도면을 참조하여 아래에서 상세히 설명된다.

도면의 간단한 설명

[0009] 본 명세서에 편입되어 그 일부분을 형성하는 수반 도면은 본 발명의 실시형태를 예시하고, 그 설명과 함께, 본 발명의 원리를 설명하는 역할을 한다. 도면에서:

도 1은 일례의 실시형태에 대한 시스템 선도의 예시도,

도 2a는 일례의 실시형태에 따른 플러그-인 프레임워크 및 아키텍처의 예시도,

도 2b는 작업 공간의 일례의 XML 기술의 예시도,

도 2c는 플러그인의 일례의 XML 기술의 예시도,

도 3은 플러그-인 프레임워크 및 아키텍처를 사용하여 원격 컴퓨터와 인터페이싱하기 위한 구현의 예시도,

도 4는 플러그-인 프레임워크 및 아키텍처를 사용하여 외부 알고리즘과 인터페이싱하기 위한 구현의 예시도,

도 5a는 일례의 실시형태에 따른 결과 피드백을 갖는 데이터 발견 노드(data discovery node) 프로세스의 하이

레벨 표현의 예시도,

도 5b는 데이터 발견 노드가 지식 베이스를 확장하도록 어떻게 사용될 수 있는지의 일례의 예시도,

도 6은 데이터 발견 노드에 의해 수행된 데이터 분석 흐름에 대한 전체 라이프-사이클의 예시도,

도 7은 데이터 발견 노드를 생성하고 연산 변수를 설정 및 정의하도록 사용된 사용자 인터페이스의 예시도,

도 8a는 전문가가 데이터 발견 노드를 훈련시키고 그리고 분석가가 전문적으로 훈련된 데이터 발견 노드를 인보크하는 것의 예시도,

도 8b 및 도 8c는 데이터 발견 노드의 전문가 훈련의 일례도, 및

도 9는 데이터 발견 노드에 의해 표현된 결정 트리의 예시도.

발명을 실시하기 위한 구체적인 내용

[0010] 유사한 참조 번호가 유사한 요소를 표시하는 수반 도면을 참조하면, 도 1은 시스템 선도를 예시한다. 도 1에서 도시된 바와 같이, 데이터 획득 기기는 획득 컴퓨터에 접속된다. 일례의 실시형태에서, 획득 기기는 흐름 세포 측정기이다. 그렇지만, 흐름 세포 측정기 이외의 기기가 획득 기기로서 사용될 수 있음을 이해해야 한다. 그렇지만, 발명자는 여기에서 설명되는 기술이 흐름 세포 측정을 포함하는 단일 세포 기술에 관하여 특히 혁신적이고 유용하다고 생각하므로, 설명의 목적으로, 흐름 세포 측정이 여기에서의 일례의 실시형태로서 사용될 것이다.

[0011] 분석 컴퓨터는, 인터넷을 통해, 서브넷을 통해, 인트라넷을 통해서와 같이, 네트워크 접속을 통해 서버에, 또는 인터넷을 통해 클라우드에 접속된다. 일부 실시형태에서, 획득 기기는 획득 컴퓨터에 접속될 수 있고, 그리고 획득 기기에 의해 획득된 데이터는 데이터를 분석 컴퓨터에 전달한 후에 분석 컴퓨터 상에서 분석된다.

[0012] 분석 컴퓨터는 분석 소프트웨어를 실행하고, 그리고 분석 소프트웨어는 시험되고 있는 샘플에 대해 획득 기기의 하나 이상의 파라미터(예컨대, 전압, 흐름율 등)를 조절할 수 있다. 그러한 분석 소프트웨어는 또한 파라미터가 올바르게 설정되는지 사용자가 평가하기 위한 피드백을 제공하기 위해 샘플 데이터를 획득하면서 초기 샘플 정보를 디스플레이할 수 있다. 분석 소프트웨어는 획득 기기의 제조자에 종속하여 달라질 수 있다. 일부 실시형태에서, 획득 컴퓨터는 사용자-인터페이스 항목을 대부분 포함하고 있는 가벼운 버전의 분석 소프트웨어를 실행할 수 있고, 그리고 서버는 또한 소정 버전의 분석 소프트웨어를 포함한다. 이러한 실시형태에서, 서버는, 서버가 획득 컴퓨터보다 더 많은 컴퓨팅 자원을 가지고 있을 수 있기 때문에, 무거운 데이터 분석과 같은 프로세싱-집약적 기능을 수행할 수 있다.

[0013] 분석 소프트웨어는 획득 기기에 의해 분석되고 있는 샘플의 결과를 표시하는 데이터 신호를 획득 기기로부터 수신할 수 있거나, 또는 분석 소프트웨어는 획득 기기에 의해 수집된 데이터를 표현하는 데이터 파일을 수신할 수 있다. 일부 실시형태에서(예컨대, 획득 기기가 흐름 세포 측정기일 때), 분석 소프트웨어에 의해 발생된 데이터는 샘플에서의 세포 수, 말초 혈액 단핵 세포(PBMC)의 수 및 빈도, CD4+ T 세포 수, CD14 세포 수, CD7+ 세포 수 등 중 어느 하나 또는 전부를 표시할 수 있다. 샘플 분석의 결과는 하나 이상의 흐름 세포 측정 표준 포맷 파일(예컨대, FCS 또는 CSV 파일) 내에 포함되어 있을 수 있다. 획득 컴퓨터는 획득 기기에 의해 제공된 신호 및 데이터에 기반하여 FCS 파일을 생성한다. 그렇지만, 특히 획득 기기가 흐름 세포 측정기가 아니면, 다른 파일 포맷이 사용될 수 있음을 이해해야 한다. 분석 소프트웨어는 획득 기기 ID, 환자 ID, 획득 조건 및 파라미터 등과 같은 것들을 표시하는 샘플에 대한 메타데이터를 더 발생시킬 수 있다.

[0014] 분석 컴퓨터는 또한 분석 컴퓨터가, 분석 서버 또는 제3자 서버와 같은, 원격 컴퓨터와 통신할 수 있게 하는 인터페이스를 포함한다. 획득된 데이터가 전달되는 다른 컴퓨터의 일례로서, 서버는 흐름 세포 측정 분석에 전용인 원격 서버일 수 있다. 원격 서버 실시형태에서, 분석 또는 획득 컴퓨터는 네트워크를 통하여 서버에 액세스할 수 있다. 분석 또는 획득 컴퓨터는 또한 제3자 컴퓨터 시스템 또는 서버와 통신할 수 있다. 분석 또는 획득 컴퓨터는, 임상 목적으로 추적 식별 번호를 포함하여, 집단을 식별하도록 구성된 알고리즘과 같은 제3자 알고리즘, 또는 획득 컴퓨터에 의해 발생된 데이터를 프로세싱하거나 데이터를 분석할 수 있는 어느 다른 외부 알고리즘이라도 저장 및 실행할 수 있다. 도 1은 분석 또는 획득 컴퓨터 시스템이 제3자 알고리즘을 저장 및 실행하는 상황을 예시하고 있기는 하지만, 서버와 같은 원격 컴퓨터 또한 제3자, 또는 "외부" 알고리즘을 실행할 수 있음을 이해해야 한다. 획득 컴퓨터는 획득 컴퓨터에 의해 수행되는 분석 및 필요성에 종속하여 다수의 원격 컴퓨터 시스템과 통신할 수 있다.

[0015] 서버는 프로세서 및 메모리는 물론, 데이터베이스와 같은, 데이터 저장소 또한 포함한다. (메모리와 같은) 비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하는 프로세서-실행가능한 명령어는 여기에서 설명된 태스크를 수행하도록 프로세서에 의해 실행될 수 있다. 데이터베이스는, 여기에서 설명되는, 데이터 발견 노드 데이터 구조를 저장할 수 있다. 유사하게, 획득 컴퓨터는 프로세서 및 메모리를 포함할 수 있고, 그리고 (획득 컴퓨터의 메모리와 같은) 비-일시적 컴퓨터-판독가능한 저장 매체 상에 상주하는 프로세서-실행가능한 명령어는 획득 컴퓨터에 대해 여기에서 설명된 태스크를 수행하도록 획득 컴퓨터의 프로세서에 의해 실행될 수 있다.

[0016] 이하의 설명은, 국한되는 것은 아니지만, (1) 외부 소프트웨어 알고리즘을 인보크 및 동화하기 위한 플러그-인 프레임워크 및 인터페이스, 및 (2) 데이터-발견 노드를 사용하는 데이터-구동식 발견 프로세스를 포함하여 여기에서 설명된 본 발명의 기술의 여러 다른 태양에 관해 상술할 것이다.

[0017] ***알고리즘 플러그-인 프레임워크 및 인터페이스***

[0018] 단일 세포 검정의 연구 내에서, 과학자 및 알고리즘 작성자는 획득 기기에 의해 수집된 데이터의 분석을 능률화하는 유용한 분석 알고리즘을 계속 발생시킨다. 예컨대, 일부 외부 분석 알고리즘은 세포 집단을 식별하도록 구성된다.

[0019] 관용적으로, 세포 집단 식별은 소위 게이팅 프로세스를 통해 수동으로 행해진다. 수동 게이팅은 일반적으로 사용자가 세포 집단을 식별시키도록 데이터 포인트 세트(클러스터) 주위에 원형 또는 다각형과 같은 형상을 수동으로 그리는 것을 수반한다. 그렇지만, 생명 과학 데이터 분석에서의 진전은 세포 집단을 식별할 수 있는 자동 게이팅 프로그램을 발생시켰다. 더욱, 세포 집단 식별 또는 어느 다른 데이터 분석 단계에 대한 컴퓨터 프로세서의 사용은 프로세서-실행된 알고리즘이 인간에 의해 수행된 수동 분석보다 더 빠르고 그리고 더 객관적으로 세포 집단을 식별하거나 다른 분석을 수행할 수 있기 때문에 어떠한 인간-생성된 병목 또는 편견도 제거할 수 있다. 집단 식별 알고리즘이 일례로서 주어지기는 하였지만, SPADE, FlowMeans와 같은 높은-처리율 유전자형 및 표현형 데이터 분석 및 분석 결과를 시각화하거나 보고서를 발생시키기 위한 외부 알고리즘, 및 Bioconductor 프로젝트의 일부분으로서 호스팅된 알고리즘과 같이, 과학자가 획득 기기에 의해 수집된 데이터를 분석 및 해석하는 것을 돕는 다른 유형의 데이터 분석 알고리즘이 존재한다.

[0020] 집단 식별을 위한 외부 알고리즘에 부가하여, 알고리즘 플러그-인 프레임워크 및 인터페이스는 오픈-소스 데이터베이스로부터 실험 데이터를 다운로드하거나, 외부 데이터베이스로부터 주석 달린 실험 데이터를 다운로드하거나, 외부 서버 또는 원격 컴퓨터 시스템이 스캔하여 통계 값을 찾을 수 있도록 작업 공간 데이터를 업로드하거나, 애플리케이션 레벨 연산을 실행하거나, 또는 임상 시험을 위한 추적 식별 번호를 수신하도록 외부 서버 또는 원격 컴퓨터 시스템과 통신할 수 있다. 외부 서버 시스템과 상호작용할 수 있는 능력은 분석 결과의 가치 있는 프리- 및 포스트-프로세싱을 분석 소프트웨어에 제공한다. 예컨대, 임상 시험을 수행하는 과학자가 시험 식별 번호가 필요하면, 알고리즘 플러그-인 프레임워크 및 인터페이스는 검증 목적으로 임상 시험 실험 결과를 업로드하도록 외부 서버와 통신할 수 있다.

[0021] 또 다른 실시형태에서, 분석 소프트웨어 내부에 있는 알고리즘은 특정 플랫폼에서 구획화되어, 그들의 의도된 컨텍스트 밖에서는 그것들에 액세스 가능하지 않게 할 수 있다. 이들 내부의, 그러나 그들의 의도된 컨텍스트 밖에서는 액세스 가능하지 않은, 알고리즘의 예는 (분석 소프트웨어가 FlowJo일 때) 증식 플랫폼에서의 다항식 핏, FlowJo의 보상 편집기에서의 +/- 피크 찾기, 또는 FlowJo의 세포 사이클 플랫폼에서의 가우시안 핏팅을 포함할 수 있다. 여기에서 설명된 알고리즘 플러그-인 프레임워크 및 인터페이스는 외부 알고리즘을 분석 소프트웨어에 통합할 뿐만 아니라 또한 위에서 설명된 그들의 현재, 제한된 컨텍스트 밖에서 구획화된 내부 알고리즘의 사용을 가능하게 한다.

[0022] 플러그인 시스템은 외부 알고리즘이 그 기능성을 확장하여 제품에서 실행되는 것을 가능하게 하기 위한 API를 제공하는 메커니즘이다. 외부 알고리즘은 전형적으로는 결과적 CLR/CSV 파일(이 경우 각각의 로우는 샘플에서의 이벤트에 대응함)을 발생시킴으로써 집단을 식별하도록 사용될 수 있지만, 보고서 또는 표와 같은 부가적 아티팩트를 또한 발생시킬 수 있다. 예시적 실시형태에서, 외부 알고리즘은 자바 언어, 또는 자바로부터 인보크될 수 있는 어느 다른 언어로라도 구현될 수 있다. 외부 알고리즘을 부가하기 위해, 개발자는, 후에 그래프 및 통계를 생성하도록 FlowJo의 기하학적으로-게이팅된 집단 노드처럼 조작될 수 있는, 작업 공간에서의 새로운 '집단 노드'를 생성하도록 FlowJo 제품에 의해 사용되는 자바 인터페이스를 구현할 것이다.

[0023] 도 1에서 도시된 바와 같이, 획득 컴퓨터는 획득 기기에 의해 획득된 데이터의 분석에 유용한 복수의 소프트웨어 프로그램 및 알고리즘을 저장 및 실행할 수 있다. 예컨대, 분석 소프트웨어는, FlowJo와 같은, 단일 세포 분

석 프로그램을 포함할 수 있다. 제3자 알고리즘은, 국한되는 것은 아니지만, 위에서 설명된 외부 서버 함수 또는 자동 집단 식별 프로그램과 같이, 분석 소프트웨어에 상보적인 프로세싱을 수행할 수 있다. 획득 컴퓨터는 분석 소프트웨어의 지시로 외부 알고리즘을 실행할 수 있다. 일부 실시형태에서는, 획득 컴퓨터가 외부 알고리즘을 실행할 수 있고, 다른 실시형태에서는, 도 1에서 도시된 서버와 같은 원격 컴퓨터가 외부 알고리즘을 실행하고 네트워크를 통하여 획득 컴퓨터에 외부 알고리즘의 프로세싱의 결과를 제공할 수 있다.

[0024] 도 2는 분석 소프트웨어의 프로세싱의 세션 내 분석 결과의 프리/포스트-프로세싱 또는 외부 알고리즘 인보크를 위한 일례의 프레임워크 및 인터페이스를 예시한다. 여기에서 설명된 프레임워크는 기존 과학적 데이터 분석 소프트웨어 상에서 구축될 수 있다. 예컨대, 분석 소프트웨어가 흐름 세포 측정 데이터를 분석하도록 발생된 소프트웨어이면, 프레임워크는 흐름 세포 측정기가 모은 데이터 내에서 세포 집단을 식별하기 위한 외부 알고리즘을 호출해 낼 수 있다. 외부 서버 및 외부 알고리즘과 상호작용하기 위한 프레임워크는 데이터 분석 소프트웨어 내에 포함될 수 있다.

[0025] 예컨대, 프레임워크는, 자바와 같은 프로그래밍 언어에 의해 정의된 바와 같은, 협업 클래스 세트 및 그들 상호작용 시퀀스를 포함할 수 있다. 일례의 프로그래밍 언어로서 자바가 주어지기는 하지만, 어느 수의 프로그래밍 언어 중 하나라도 여기에서 설명된 프로세스 및 프레임워크를 실행하는 프로그래밍 언어로서 역할을 할 수 있다. 다수의 프로그래밍 언어가 여기에서 설명된 시스템 및 방법을 달성할 수 있기는 하지만, 자바는 그것이 다른 프로그래밍 언어에 비해 바람직하게 하는 특정 이점, 즉, C, R 또는 웹-기반 계산 엔진 언어와 같은 다른 프로그래밍 언어를 호출해 낼 수 있는 자바의 능력을 가지고 있다. 과학적 기기에 의해 수집된 데이터의 통계적 분석을 수행하는 많은 외부 알고리즘은 R 언어로 작성된다. 그리하여, R을 호출해 낼 수 있는 자바의 능력은 R로 작성된 외부 알고리즘과 분석 소프트웨어를 가교한다. 물론, 외부 알고리즘이 R로 작성되어 있지 않으면, 자바는 또한 외부 알고리즘의 프로그래밍 언어를 호출해낼 수 있다.

[0026] 프레임워크는 현재 및 장래 데이터 분석 알고리즘이 데이터 값 입력 세트로 인보크되게 하는 메커니즘은 물론, 또한, 이벤트 클러스터 값, 공식, 시각적 그래픽, 또는 기하학적으로-정의된 경계 정의 형태의, 분석 결과의 후속 프로세싱을 제공한다. 환언하면, 프레임워크는 입력 데이터 세트를 발생시키고 그리고 입력 데이터를 외부 알고리즘 또는 외부 서버에 통신하기 위해 2개의 인터페이스 중 하나를 호출해 낸다. 외부 알고리즘의 프로세싱 후에, 프레임워크는 외부 알고리즘 또는 서버로부터 분석 결과를 수신하고 그리고 알고리즘 또는 프리/포스트 프로세싱의 인보크가 표현되고 파일에 저장되게 하는 메커니즘을 제공한다. 파일에 저장된 분석 결과는 다운스트림 통계적 계산, 결과의 그래프 작성, 또는 (부가적 외부 알고리즘, 후속 프리/포스트-프로세싱, 또는 분석 소프트웨어 내에 포함된 알고리즘과 같은) 다른 알고리즘의 인보크를 위해 분석 소프트웨어와 통합될 수 있다.

[0027] 프레임워크는 또한, 데이터 분석 소프트웨어 자체의 외부에 있는 알고리즘인, 통합된 알고리즘의 인보크를 관리한다. 분석 소프트웨어는 생물학자가 이들 알고리즘과 상호작용할 수 있는 인터페이스를 제공한다. 분석 소프트웨어는, 생물학자(예컨대, 분석이 실행되어야 하는 특정 집단을 선택함) 및 플러그인 개발자(예컨대, 알고리즘이 입력(예컨대, 생물학자가 선택한 집단의 데이터 값에 대응하는 CSV 파일)으로서 필요로 하는 데이터에 대한 요건 및, 분석 다음에, 어디서 그리고 무슨 유형의 출력이 플러그인 인터페이스가 사용자에게 제시하는데 이용가능할지를 특정함) 양자에 의해 제공된 명령어에 기반한다. 인터페이스는 또한, 분석이 항상 계층적으로 올바르고 생물학적으로 관련 있게 머물게 되도록, 분석에서의 업데이트가 통해 통신되는 에이전트로서 역할을 한다. 더 구체적으로, 프레임워크는 분석이 처음 실행될 때 통합된 알고리즘을 인보크할 뿐만 아니라, 프레임워크는 또한 데이터 값 입력 세트가 변경될 때마다 통합된 알고리즘을 재-실행한다. 그래서, 과학자는 다수의 데이터 입력 세트에 대해 신속히 분석을 실행할 수 있고, 그리고 프레임워크는 입력 데이터 값이 변경되거나 사용자가 실험 파라미터를 변경할 때는 언제든지 사용자 상호작용 없이 통합된 알고리즘을 인보크 및 재-실행할 수 있다. 예컨대, 일부 데이터 파라미터를 변경하는 것은 통합된 알고리즘에 의해 어떻게 집단이 식별되는지를 변경할 수 있다. 데이터 입력에서의 변경을 알아차릴 때, 프레임워크는 집단을 재-식별하기 위해 통합된 알고리즘을 인보크하고, 그리고 프레임워크는 통합된 알고리즘에 의해 발생된 분석 결과를 사용한다. 통합된 알고리즘으로부터 분석 결과를 수신시, 프레임워크는 분석 소프트웨어에 의해 이해되는 데이터 포맷으로 분석 소프트웨어에 결과를 제공할 수 있고, 그리고 분석 소프트웨어는, 통계적 분석, 그래프 작성, 또는 보고와 같은, 결과에 대한 다운스트림 분석을 수행할 수 있다.

[0028] 프레임워크는 작업 공간이 장래 분석을 위해 저장되고 다시 열릴 수 있도록 알고리즘 통합이 작업 공간으로서 저장될 수 있게 한다.

[0029] 프레임워크는 원격 컴퓨터 시스템과 통신하기 위한 인터페이스 및 외부 알고리즘과 통신하기 위한 인터페이스를

포함한다. 각각의 인터페이스는 외부 서버 상에 저장된 함수 또는 외부 알고리즘이 사용자 상호작용 없이 인보크될 수 있게 하는 수단을 제공한다. 사실, 그래픽 사용자 인터페이스를 통해 데이터 프로세싱을 보고 있는 사용자에게는, 외부 알고리즘의 인보크는 보이지 않는데, 통계, 그래프, 또는 분석 소프트웨어에 의해 발생된 다른 보고서를 통해서와 같이, 외부 알고리즘에 의해 수행된 분석의 결과만이 사용자에게 보여질 수 있기 때문이다.

[0030] 일반적으로, 통합된 알고리즘의 인보크를 위한 인터페이스는, 국한되는 것은 아니지만, 데이터 값의 입력 파일, 출력 폴더 목적지, 및 하나 또는 다수의 실험으로부터의 데이터 세트의 XML 기술을 포함한다. 이러한 XML 기술은 원시 데이터로의 포인터, 플러그인-구동식 분석을 포함하는 실행된 모든 분석, 데이터에 대한 메타-정보, 및 로지클, 바이익스포넨셜, 하이퍼로그, 및 하이퍼볼릭 아크사인과 같은 데이터를 프로세싱 및 시각화하는데 최적으로 사용되는 데이터 변환을 포함할 수 있다. XML 기술은 이러한 정보를 마크업을 통하여 특정하여 원시 데이터를 분석 및 연관된 결과에 계층적으로 링크하는 XML 문서의 형태를 취할 수 있다. 도 2b는 작업 공간의 일례의 XML 기술을 도시하고, 그리고 도 2c는 플러그인의 일례의 XML 기술을 도시한다. 동일한 데이터 및 분석 아키텍처를 저장할 수 있는 전유 2진 파일과 같은, XML 이외의 형태가 사용될 수 있음을 이해해야 한다. 더욱, 데이터 세트의 기술은, XML로든 다른 포맷으로든, 외부 알고리즘에 의해 산출된 어느 유도 데이터에 대해서라도 포인터 및 어느 플러그인-기반 분석에 대해서라도 입력 파라미터에 관한 메타데이터를 포함할 수 있다. XML 메타-정보가 외부 알고리즘에 의해 사용되는지는 인보크된 알고리즘에 종속한다. 외부 알고리즘 인터페이스는 또한 알고리즘 인보크가 저장되고 추후 프레임워크에 의해 복원되는 단계를 정의한다. 인터페이스는 통합된 알고리즘으로부터 그래픽, 유도된 파라미터, 표 데이터, (게이팅 ML 포맷에서와 같은) 게이팅 데이터, CLR(classification result) 파일, XML 데이터, 또는 CSV(comma separated value) 파일의 형태로 분석 결과를 수신할 수 있다. 다르게 말하면, 인터페이스는 통합된 알고리즘에 의해 발생된 아티팩트를 관리하도록 구성된다.

[0031] 인터페이스는 외부 알고리즘 및 서버 함수가 외부 알고리즘을 분석 소프트웨어에 플러그인 하도록 고수해야 하게 하는 계약을 정의한다. 외부 알고리즘 인터페이스 및 프리/포스트 프로세싱 인터페이스는 외부 서버 상의 프리/포스트 프로세싱과 인터페이싱하거나 외부 알고리즘과 인터페이싱하기 위한 계약을 각각 정의한다. 다른 인터페이스 구현 단계는 도 3 및 도 4에서 더 상세히 예시된다.

[0032] 도 3을 참조하면, 원격 컴퓨터와 인터페이싱하기 위한 구현 단계가 예시된다. 방법은 인터페이스가 작업 공간을 여는 것으로 시작한다. 작업 공간을 여는 것은 프로세서가 작업 공간의 XML 및 프리/포스트-프로세싱 인터페이스의 XML을 판독하는 것을 포함한다. 작업 공간 XML은 각각의 샘플과 연관된 메타데이터(획득된 날짜, 기기 유형, 파라미터 이름 등)은 물론 또한 획득-후 부가된 어느 사용자-정의된, 샘플-특정적 메타데이터라도 포함하고 있는 한편, 플러그-인 인터페이스에 특정적인 XML은 플러그인 모듈의 실행/업데이트에 필요한 변수, 예컨대, 데이터베이스 또는 서버의 URI를 보유한다. 작업 공간을 판독하고 URI를 수신하는 결과로서, 프로세서는 서버 또는 거기에 저장된 데이터 스토어(예컨대, 데이터베이스)로의 접속을 확립하여 아래에 설명된 바와 같이 인증을 개시하고, 쿼리를 실행하고, 데이터베이스로부터의 데이터를 검색하고 작업 공간 XML을 수정한다. 작업 공간 열기 단계는 프로세서에 의해 실행된 프리/포스트-프로세싱 인터페이스가 데이터베이스(예컨대, 특정된 데이터 파일에 대한 분석 명령어 및 메타데이터를 포함하는 샘플 추적을 위한 LIMS(Laboratory Information Management System))로부터의 검색에 기반하여 (분석 소프트웨어에 입력된) XML 작업 공간을 증강 또는 수정하는 것을 더 포함한다. 부가적으로, XML 입력은 게이트, 통계, 샘플 이름, 또는 작업 공간 XML에 포함되어 있을 수 있는 어느 것이라도 부가하도록 수정될 수 있다. 입력이 분석 소프트웨어에 의해 정의된 정의된 스키마를 고수하는 한, 이들 부가는 분석 소프트웨어에서 계산 및 표현을 인보크할 수 있다. 입력의 유효성 검증 및 적절한 오류 보고는 인터페이스를 통해 취급되고, 그리고 입력을 시험하기 위한 유효성 검증조가 전개시 실행된다. 그것은 또한 권한 부여를 수행할 수 있는데, 분석 소프트웨어가 서버로의 액세스를 가짐을 확실히 하는 것, 외부 서버가 온라인인지를 결정하는 것, 크리덴셜을 교환하는 것, 또는 어느 다른 권한 부여 단계의 형태로라도 올 수 있다. XML 증강은 프리/포스트-프로세싱 단계가 원격 서버에 의해 수행되어야 함을 반영하도록 프로세서가 메타데이터를 발생 또는 변경하는 것을 포함한다.

[0033] 다음에, 방법은 분석 소프트웨어 내에 작업 공간을 저장한다. 저장 단계는 프로세서가 작업 공간 및 프리/포스트 프로세싱 인터페이스의 상태를 저장하는 것을 포함한다. 플러그인은 작업 공간에서 그 자신의 XML 표현을 업데이트하여 그 '상태'를 보유할 것이고 그리고/또는 XML을 순회하여 데이터를 추출하고 그리고 액션, 예컨대, 데이터베이스를 특정된 통계로 업데이트하는 것을 수행할 수 있다. 이러한 단계 동안, 프리/포스트-프로세싱 인터페이스는 취해진 분석 액션의 로그 또는 SQL 출력과 같은 부가적 아티팩트를 발생시킬 수 있고, 그리고 프리/

포스트-프로세싱 인터페이스는 외부 시스템과 통신한다. 이러한 통신 동안, 인터페이스는 외부 시스템에 입력 데이터를 제공하고 그리고, MIFlowCyt 표준에 따라 수집 및 주석 달린 데이터를 다운로드하는 것, 임상 추적기로부터 추적 식별 번호를 수신하는 것, 또는 어느 다른 프리/포스트 프로세싱 단계와 같이, 외부 시스템으로부터 데이터를 수신한다. 프리/포스트 프로세싱 인터페이스는 이러한 통신을 하기 위해 서버 URL을 참조로 할 수 있다.

[0034] 외부 서버와의 통신을 완료한 후에, 프로세서는 세션을 종료하고, 그리고 프리/포스트 프로세싱 인터페이스는 데이터베이스 접속과 같은 컴퓨터 자원을 풀어준다.

[0035] 이제, 도 4를 참조하면, 외부 알고리즘과 인터페이싱하기 위한 구현 단계가 예시된다. 방법은, 외부 알고리즘 인터페이스를 위한 랩퍼로서 정의될 수 있는, 외부 집단 노드를 생성함으로써 시작한다. 이러한 생성 단계 동안, 프로세서는, 파라미터를 설정하는 것, 연산 변수를 설정하는 것, 파일 명명 등과 같이, 외부 알고리즘과 관련 있는 옵션을 사용자에게 프롬프트할 수 있지만, 이러한 사용자 프롬프트 단계는 옵션이고 그리고 인보크된 외부 알고리즘에 종속할 수 있다.

[0036] 다음으로, 프로세서는 외부 알고리즘에 의해 수행되는 계산을 인보크하기 위해 XML 표현을 발생시킴으로써 엔진 요청을 구성한다. XML 표현은 무슨 알고리즘을 실행할지 또는 발생시킬 시각화, 및 연관된 입력 및 필요한 인수, 예컨대, 파일 경로, 파라미터의 수, 클러스터의 수, 차원 축소를 위한 변수, 색 선택, 시각화의 유형, 저장을 위한 이미지 유형 등을 표현한다.

[0037] 요청을 구성한 후에, 프로세서는 외부 알고리즘을 인보크한다. 외부 알고리즘을 인보크하는 것은 FCS 파일, FCS 파일과 포함된 XML(이벤트의 수, 샘플 파일 이름, 및 집단 이름을 포함함), 및 외부 알고리즘이 그 결과를 저장하여야 하는 출력 폴더를 외부 알고리즘에 제공하는 것을 포함한다. 응답하여, 외부 알고리즘은 그 프로세싱 및 계산을 수행한다. 외부 알고리즘이 요청된 프로세싱 및 계산을 수행한 후에, 분석 소프트웨어 인터페이스는 결과를 수신하고 그것들을 분석 소프트웨어에 통합한다. 이들 결과는 CSV 파일, CLR 파일, 게이팅ML 파일, 또는 FCS 파일의 형태로 올 수 있다. CSV 또는 CLR 파일을 들여올 때, CSV 또는 CLR 파일의 각각의 로우는 FCS 파일에서의 이벤트에 대응하고 그리고 칼럼 수는 클러스터 수에 대응한다. 더욱, 외부 알고리즘 인터페이스는 유도된 파라미터를 생성하고, 그리고 분석 소프트웨어는 유도된 파라미터에 관하여 자동으로 게이팅하여 서브-집단을 생성한다. 결과를 수신한 후에, 프로세서는 알고리즘으로의 입력을 수정할 수 있다. 일 실시형태에서, 프로세서는 주어진 출력 파일에 저장된 데이터를 참조로 함으로써 외부 알고리즘의 결과를 수신한다.

[0038] 외부 알고리즘으로부터의 결과를 수신한 후에, 프로세서는 파일 시스템에 작업 공간을 저장하고 분석 소프트웨어 작업 공간을 복원한다. 프로세서는 그 후 분석 소프트웨어의 지시로 부가적 다운스트림 분석을 수행할 수 있다.

[0039] 이러한 식으로, 외부 서버 상에 저장된 함수 및 외부 알고리즘은 분석 소프트웨어에 완전 통합 없이 분석 소프트웨어에 이용가능하다. 분석 소프트웨어의 사용자는 주요 작업 흐름 해킹 또는 커맨드 라인 지식 없이 무수한 더 많은 분석 옵션 및 기능성을 얻는다. 대신, 사용자는 서버 상에 저장된 외부 함수 또는 외부 알고리즘을 무결절 인보크하도록 분석 소프트웨어의 그래픽 사용자 인터페이스를 사용할 수 있다.

[0040] ***데이터 발견 노드 아키텍처 및 프로세스***

[0041] 분석 소프트웨어 내에서, "노드"는, 기하-기반 툴을 사용하여 기하학적 클러스터를 정의하거나 획득 기기에 의해 획득된 데이터에 통계적 분석을 적용하는 단계와 같은, 전체 분석 단계를 표현한다. 그러한 "노드"는 입력, 전 세트 또는 서브세트 또는 이벤트-레벨 원시 데이터, 및 수학적 모델(예컨대, 세포 사이클에서 세포의 백분율), 또는 세포 서브세트의 기하학적 정의와 같은, 출력을 갖는 계산 또는 프로세싱 단계를 표현한다. 환언하면, 노드는 분석 소프트웨어가 집단 식별, 통계적 계산, 수학적 함수, 기하학적 게이팅, 결과 제시, 결과 증강 등과 같은 분석 계산을 수행하라고 분석 소프트웨어에 명령하는 것에 의해 생성된 데이터 구조이다. 부가적으로, 노드 데이터 구조는, CSV 파일, 게이팅ML 파일 등에서와 같이, 결과를 제시할 방식 및 분석 함수에 입력할 데이터의 규격을 포함한다. 데이터 구조는 더욱 데이터 입력의 유형 나름일 수 있다.

[0042] 여기에서 설명된 기술은 사용자가 데이터 분석 애플리케이션 내에서 "데이터 발견 노드"(DDN) 프레임워크를 통해 데이터 세트에 관하여 데이터 분석을 특정 및 수행할 수 있도록 위에서 설명된 노드 개념을 확장하는데, 여기서 DDN 프레임워크는 데이터 분석 소프트웨어 자체에 어떤 지능이 이미 상주할 수 있었든 그 너머로 광범위한 지식 베이스로의 액세스를 갖는 데이터 분석을 제공한다. 예컨대, DDN는 또한 위에서 개시된 플러그-인 인터페이스 및 프레임워크를 사용하여 분석 소프트웨어에 플러그인된 외부 알고리즘으로부터 이루어질 수 있는 결정을

캡슐화할 수 있다. 알고리즘-기반 결정은 의사-결정을, 주관적 편견을 갖는 개개의 분석가로부터 떠나, 데이터-구동식 알고리즘으로 옮김으로써 분석의 주관성을 제거한다. 여기에서 설명된 데이터 발견 노드 아키텍처 및 프로세스는 또한 단방향 노드를 적어도 다음의 4개의 목표를 성취하는 능동 노드로 변환한다: 1) 능동 노드는 샘플, 그룹, 및 연구 간 비교를 제공하도록 반복된, 재현가능한 분석이 가능하다(즉, 분석가의 주관적 편견에 의해 영향을 받지 않음); 2) 능동 노드는 드래그-앤-드롭 메커니즘을 통해 복잡한 분석 및 보고에 대한 장벽을 낮춘다; 3) 능동 노드는 입력 데이터가 변경되면 업데이트를 하기 위해 라이브인 채로 있다; 그리고 4) 능동 노드는 노드가 분석에서 스태킹되고 커맨드 라인 모드로 실행될 수 있으므로 자동화를 용이하게 한다.

[0043] 도 5a는 능동 노드 아키텍처를 표현한다. 일례의 실시형태에 따른 DDN 프레임워크는 이하의 컴포넌트를 포함한다: 소프트웨어, 데이터 구조, 알고리즘, 및 네트워크를 통하여 액세스가능한 데이터베이스. 위에서 언급된 바와 같이, 데이터 분석 애플리케이션에서의 DDN 프레임워크는 기본적으로는 사용자가 노드의 각각의 적용시 구축되는 지식 베이스에 액세스하는 방식이다. 그리하여, 사용자는 완료될 분석에 대한 DDN를 생성하도록 사용자 인터페이스를 통하여 제스처를 취하고, 그리고 무슨 유형의 분석이 완료되어야 하는지 표시한다. (작업 공간에서 "노드"로서 물리적으로 발현되는) DDN의 최종-사용자 인스턴스는 지식을 포함하고 있지 않고 그보다는 그것은 사용자가 그들이 분석하고 있는 것(예컨대, 노인에서 참조 CD3+ 백분율)의 더 큰 컨텍스트에 플러그인할 수 있게 한다. 예컨대,

[0044] a. 사용자는 게이트 노드로서 T 세포 집단을 식별하도록 CD3+ 게이트를 생성한다,

[0045] b. 노드는 (로컬 클라이언트에 있는, 즉, "이것을 DDN으로 만드는") 사용자 인터페이스에서 사용자에 의해 DDN로서 배정되어 2개의 결과를 갖는다:

[0046] i. 이하의 집단 및 샘플 정보가 지식 베이스의 물리적 메모리에 기록된다:

[0047] 1. "샘플 정보"

[0048] a. FCS(원시) 파일에 포함되어 있는 메타데이터, 예컨대, 어느 기기 상에서인지, 무슨 획득 소프트웨어에 의해서인지

[0049] b. DDN 지식에 보낼 샘플 컨텍스트(세포 유형, 종)

[0050] 2. 아래에서 개괄되는, DDN 실행 파라미터

[0051] 3. 생물학적 결과 정보 - 분석의 수치 결과 및 통계

[0052] ii. DDN이 반복 n>1에 있으면, DDN는 "나의 데이터에 기반하여, 이러한 CD3+ 빈도는 이전 관찰 아래에 있는 2개의 표준 편차이다"와 같은 어느 플래그라도 사용자에게 반환한다.

[0053] c. 그리하여, 지식 베이스는 참조를 제공하고, 그리고 DDN는 당면 분석가와, 전술한 예의 DDN 파라미터를 통하여 확립된 바와 같은 현재 패턴과 매치하는 이전 분석가의 데이터 간 양방향 다이얼로그를 제공한다(도 5b 참조). DDN는 이러한 교환을 "구동하는" 컴포넌트, 즉, ("원시" 데이터 파일에 액세스하는 프로세서와는 별개의) 호스팅된 네트워크를 가능으로 하는 사용자-대면 노드이다.

[0054] 도 5a 및 도 5b에 의해 도시된 바와 같이, 입력 데이터는 DDN에 제공되고, 그리고 DDN는 결과를 산출하는 분석 단계를 수행한다. DDN에 의해 발생된 결과 데이터는 DDN에 피드백될 수 있거나, 또는 결과 데이터는, 입력 데이터를 프루닝하는 것, 입력 데이터로부터 잡음을 제거하는 것, 또는 입력 데이터의 파라미터를 변경하는 것에 의해서와 같이, 입력 데이터를 변경한다. 결과 데이터가 입력 데이터에 어쨌든 영향을 미칠 때, DDN는 새로운 데이터 세트로 동일한 분석 단계를 적용할 수 있거나, 또는 DDN는 새로운 데이터 세트에 기반하여 다른 분석 단계를 적용할 수 있다 - 이러한 식으로, DDN는 제1 반복 후에 "데이터-구동된다"고 생각될 수 있다.

[0055] 더욱, 결과 데이터는 다운스트림 프로세싱과 더 관련될 수 있다. 예컨대, DDN는 집단 식별 알고리즘을 표현할 수 있고, 그리고 결과 데이터는 결정적이 아닌 또는 바람직하지 못한 결과를 산출할 수 있다. 노드는 결과 데이터를 분석할 수 있고, 그리고 결과 데이터의 분석에 기반하여, DDN는 입력 데이터 내 집단을 더 양호하게 식별하도록 집단 식별 알고리즘의 파라미터를 변경할 수 있다. 다른 일례에서, 결과 데이터는 식별된 표현형(예컨대, CD8+)이 이환율 또는 치료 효능과 상관을 가지고 있지 않다고 결정할 수 있다. 이환율 또는 치료 효능과의 상관이 결과 데이터에 의해 발견될 수 없으면, DDN 또는 DDN를 훈련시키는 과학자는 장래 분석에 대해 이러한 표현형을 무시하라고 DDN에 명령할 수 있다. 이러한 식으로, DDN는 참조된 집단 식별 알고리즘을 사용하여 집단을 가장 정확하게 식별하도록 최적화된다. 위에서의 예에 의해 알 수 있는 바와 같이, 데이터 및

알고리즘은 DDN에 의해 이루어지는 결정을 구동한다. DDN이 더 많은 데이터를 수신할수록, 그리고 DDN이 더 많이 프로세싱할수록, DDN는 더 많이 학습한다. 이러한 데이터-구동식 방법은 아래에서 더 상세히 설명될 것이다.

[0056] 또한, 실무자는 프레임워크가 공격에 덜 취약하도록 DDN 프레임워크에 보안 또는 큐레이션 층을 포함시키기로 선택할 수 있음을 유념해야 한다. 이것은 불량한 또는 훈련되지 않은 행위자가 지식 베이스를 더럽히는 것(예컨대, 100명의 사람들이 1%의 잘못된 CD3+ 빈도를 게이팅하고 그 불량 데이터를 제출하는 것)을 방지하는 것을 도울 수 있다.

[0057] 도 6은 DDN에 의해 수행되는 데이터 분석 흐름에 대한 라이프-사이클을 예시한다. 도 6에 의해 예시된 프로세스에서, 직사각형으로서 예시된 방법 단계는 프로세서에 의해 수행된 액션 단계를 표현하고, 다이아몬드로서 예시된 방법 단계는 프로세서에 의해 수행된 결정 단계 또는 시험 단계를 표현하고, 그리고 타원형에 의해 표현된 방법 단계는 시험 단계의 가능한 결과를 표현한다.

[0058] 하이 레벨에서, 도 6에 표현된 방법은 3개의 위상을 포함한다: 제1 DDN 사이클, 다음의 n DDN 사이클 위상, 및 완료 위상. 제1 DDN 사이클 위상은 한 번만 수행되는 반면, 다음의 n DDN 사이클은 충족 기준이 만족될 때까지 계속 반복할 수 있다. 방법은 충족 기준이 만족된 후에만 완료 위상에 들어갈 것이다.

[0059] 이제, 도 6에서 묘사된 방법이 어떻게 동작하는지 더 잘 이해하기 위해 DDN 함수를 정의 및 제어하는 데이터 객체의 유형이 설명될 것이다. 이들 데이터 객체는 연산 변수, 일시적 객체, 포인터, 메타데이터, 및 원시 리스트 모드 데이터를 포함한다.

[0060] 제1 DDN는 연산 변수 데이터 객체를 포함한다. 연산 변수는 사용자 또는 분석 소프트웨어 중 어느 것에 의해 설정된 변수로서 1) 충족 변수 임계치, 2) 메타데이터 규칙, 및 3) 특정된 데이터에 관하여 수행할 분석 소프트웨어 알고리즘 또는 연산의 규격을 포함하고 있다. 충족 변수는 사용에 의해 설정된 임계치일 수 있으며 DDN 사이클 완료를 고려하기 위해 충족되어야 한다. 메타데이터 규칙은 입력에 의해 충족되어야 하는 기준을 정의한다. 예컨대, 메타데이터 규칙은 입력 데이터가 원시 데이터의 메타데이터에서 CD4 파라미터를 나타내 보임을 특정할 수 있다. 특정된 분석 소프트웨어 알고리즘 또는 연산은 외부 알고리즘, 분석 소프트웨어 내에 포함된 수학적 함수, 또는, FlowJo의 다변량 그래프 작성, FlowJo의 보고서 발생, 기하학적 평균 발생, 집단 식별과 같은, 분석 소프트웨어 내에 포함되어 있는 어느 다른 함수, 또는 분석 소프트웨어 또는 플러그-인된 외부 알고리즘에 의해 제공된 어느 다른 함수라도 포함할 수 있다.

[0061] 도 7은 DDN를 생성하고 연산 변수를 설정 및 정의하도록 사용된 사용자 인터페이스를 예시한다. 우선, 사용자는 파일을 선택하고 그리고 발견하라는 제스처를 취한다. 파일은 획득 기기로부터 수집된 그리고 획득 컴퓨터 내 디스크 드라이브에 저장된 데이터 세트일 수 있다. 이러한 제스처는 사용자가 선택된 파일에 DDN를 적용하기를 원함을 분석 소프트웨어에 알려준다. 제스처는 사용자가 파일을 오른쪽-클릭하는 것, 키보드 단축키를 사용하는 것, 그래픽 사용자 인터페이스 내 아이콘을 클릭하는 것, 또는 프로세서에 의해 이해되는 어느 다른 제스처라도 포함할 수 있다. 발견하라는 제스처를 취한 후에, 사용자는 새로운 DDN를 훈련시키는 것 또는 데이터베이스 또는 다른 파일 저장 컨테이너에 저장된 DDN를 적용하는 것 중 어느 것을 선택할 수 있다. 사용자가 데이터베이스로부터 DDN를 다시 불러오는 것을 선택하면, 획득 컴퓨터는 DDN 데이터 구조를 저장하고 있는 데이터베이스를 호출해 내고, 저장된 DDN의 리스트를 제시하고, 그리고 사용자가 분석을 위해 DDN 중 하나를 선택할 수 있게 한다(예시되지 않음). 사용자가 새로운 DDN를 훈련시키는 것을 선택하면, 획득 컴퓨터는 DDN를 정의할 연산 변수의 리스트를, 그래픽 사용자 인터페이스를 통해, 제시한다.

[0062] 도 7은 선택을 위한 예시적 연산 변수 세트를 예시하지만, 본 발명은 도 7에서 도시된 연산 변수로 한정되지 않는다. 연산 변수는, 파라미터, 특징, 반복 변수, 및 범위 변수와 같은, 세트로 그룹화될 수 있지만, 더 많은 연산 변수 그룹이 사용자 인터페이스 내에서 정의 및 제시될 수 있다. 예컨대, 사용자는, 국한되는 것은 아니지만, 전방-산란 광(FSC), 측면-산란 광(SSC), 형광 1(f11), 형광 2(f12), 형광 3(f13), 형광 n 등과 같은 파라미터로부터 선택할 수 있다. 파라미터 선택은 단일 세포 분석에서 중요한 역할을 하고, 그리고 DDN는 그것이 적용되는 데이터의 유형에 부가하여 그 자신의 연산에 대한 메타데이터, 즉, "실행 파라미터"를 포함하고 있다. 선택된 파라미터의 예는 다음을 포함할 수 있다:

[0063] a. 세포 표현형이 정의된 파라미터. 일례로서 흐름 세포 측정을 사용할 때, 산란 파라미터는, 예컨대, 혈액에서 주요 세포 서브세트를 식별하는데 유용한, 크기 및 입상의 상대적 척도인 반면, 형광 파라미터는 생물학적 분자의 측정이다. 그리하여, 파라미터는 기본적으로는 호환가능하지 않고, 그리고 DDN 및 그 계층에 대해 선택된 레벨에서 사용된 파라미터는 분석의 재현가능성을 용이하게 하는 생물학적으로 관련 있는 정보이다.

[0064] b. 집단을 식별하도록 사용된 어느 알고리즘에 대한 유형 및 입력 변수, 예컨대, 게이팅 및 분석 정보에 관한 파라미터(정점, 인접 집단의 위치, 게이트 유형, 집단 특성(볼록한, 드문 등), 집단 이름, 게이트가 도출된 파라미터, 부모 게이트(온톨로지), 집단을 식별하도록 사용된 알고리즘).

[0065] c. 일련의 알고리즘 및 계산에 대한 계층적 분석의 유형의 수(및 그리하여 연산 순서).

[0066] 이러한 식으로, 연구자는, CD3+의 자식이고, 라이브이고, 그리고 림프구 기하학적으로 정의된 게이트인, CD4 대 CD8 형광 파라미터에 관하여 실행된 k-평균 클러스터링 알고리즘(여기서 k=3은 입력 변수이었음)에 의해 식별된, CD8+ T 세포의 집단을 특정할 수 있다. DDN는 지식 베이스로의 그리고 그로부터의 이러한 정보의 전송을 가능하게 한다.

[0067] 사용자 인터페이스는 이들 파라미터를 재명명할 수 있는 능력 역시 사용자에게 준다. 사용자는 또한 DDN에 의해 프로세싱될 데이터의 양을 제한하도록 이들 흐름 세포 측정기 미리 설정된 파라미터 중 어느 것이라도 배제할 수 있다. DDN는 분석할 파라미터, (피크, 밸리, 또는 범위와 같은) 분석할 특징, 반복할지, 및 어느 범위를 분석할지의 선택을 수신한다. 이들 및 잠재적으로 다른 연산 변수를 선택한 후에, 컴퓨터는 DDN 데이터베이스에 또한 저장될 새로운 DDN를 생성한다. 생성된 DDN는 데이터를 분석하거나, 결과를 발생시키거나, 또는 분석 소프트웨어 내에 포함되어 있는 또는 플러그-인 인터페이스 및 프레임워크를 통해 분석 소프트웨어가 액세스가능한 어느 다른 함수라도 할 준비가 되어 있다.

[0068] DDN를 설정하기 위해, 프로세서는, 이벤트 세트 또는 등가의 다소 내포적 의미를 갖는 파일 세트인, 입력 데이터의 선택을 수신한다(예컨대, 다수 시점에 걸쳐 캡처링된 CD3 측정치). 입력 데이터는 단일 샘플 또는 샘플 그룹일 수 있다. 입력 데이터를 선택한 후에, 프로세서는 입력 데이터에 종속하여 이용가능한 분석의 유형을 결정할 수 있다. DDN 데이터베이스가 설정되고 나면, 제1 단계는 "전문가"가 참조 세트를 생성하도록 실행 파라미터 및 샘플 정보 양자로 지식 베이스를 시딩하게 하는 것이다. 우리의 상기 예를 계속하면, 노인 환자로부터의 CD3+ 데이터는 전문가에 의해 정의된다. 비-전문가는 '새로운' 샘플에 관한 DDN를 생성하고, 그리고 DDN는 그것이 전문가-주도 분석을 재-생성할 수 있는지 조사하기 위해 샘플 및 실행 파라미터 양자를 비교한다. 그 매치가 존재하고 나면, 그것은 생물학적 결과 정보 - 현재 측정치 대 지식 베이스를 비교한다. 지식 베이스에 정보를 구축하는 것을 통하여 DDN의 "훈련"은 사용으로 축적되고, 그래서 지식 베이스의 CD3+ 부분으로의 각각의 쿼리는 기지의 범위의 풀에 새로운 생물학적 결과 정보를 퇴적시킨다. 이러한 2-위상 접근법은 (1) 분석이 적용 및 실행되고 (2) 참조 데이터의 지식 베이스에 비교될 수 있음을 유효성 검증한다.

[0069] 환언하면, DDN이 무엇을 계산 및 실행할 수 있는지는 입력 데이터에 종속한다. 일례에서, 프로세서는 프로세스가 선택된 데이터에 관하여 CD4 집단 식별 알고리즘을 실행할 수 있는지 결정하기 위해 CD4 이벤트가 로딩된 메타데이터에 존재하는지 결정할 수 있다.

[0070] 도 7은 파라미터로서 형광 1, 피크 특징, 고정된 카운트 반복 변수 2, 및 범위 변수에 대해 2 내지 98의 백분위수의 예시적 사용자 선택을 예시한다. 사용자가 연산 변수를 설정한 후에, 사용자 인터페이스는 생성된 데이터 발견 노드를 선택된 파일 밑에 디스플레이한다. 사용자는 장래 참조를 위해 데이터 발견 노드를 재명명할 수 있지만, 예시 목적으로, 도 7은 "발견 노드"라고 명명된 바와 같은 생성된 데이터 발견 노드를 예시할 뿐이다. 데이터 발견 노드에 대한 이들 예시적 선택은 도 7의 저부-우측 코너에 또한 예시되는 게이팅 트리와 등가이다. 그리하여, 도 7에서 도시된 예시적 연산 변수의 선택은 다음의 게이팅 트리와 같다:

[0071] • Comp - APC -Ax700 - A 서브세트. 이것은 보통은 수동으로 정의되었을 서브세트이다. 이러한 예에서, DDN는, 위에서 개괄된 그 파라미터를 통하여, 지식 베이스로부터의 정보를 사용하여 알고리즘으로 이러한 집단을 식별하고, 피크 찾기(집단 식별을 위한 다른 알고리즘 방법)를 수행하고, 그 후 자식 아집단에 통계 계산을 인보크하는데, 그 순서로이다.

[0072] ○ 피크 1

[0073] ■ 기하학적 평균: CD3(Comp - APC -Ax700 - A 서브세트). 사용자는 분석 애플리케이션 툴을 사용하여 Comp - APC -Ax700 - A 서브세트 집단의 기하학적 평균을 계산하였다, 도 7의 저부 우측에서의 선도는 이러한 분석의 계층 및 표현을 사용자에게 보여준다.

[0074] ■ 중앙값: CD3 (Comp - APC -Ax700 - A 서브세트) - 기하학적 평균에 대한 위의 것과 같지만 이 경우에서는 중앙값에 대해서이다.

[0075] ○피크 2

[0076] ■ 기하학적 평균: CD3(Comp - APC -Ax700 - A 서브세트)

[0077] ■ 중앙값: CD3 (Comp - APC -Ax700 - A 서브세트)

[0078] 도 6을 다시 참조하면, 연산 변수에 부가하여, DDN는 제1 계산 후에 일시적 데이터 객체를 발생시킨다. 일시적 데이터 객체는 DDN 진행 객체를 표현한다. 일시적 데이터 객체는 적어도 계산 반복 및 충족 변수를 포함하고 있을 수 있다. 계산 반복은 DDN에 의해 수행되는 각각의 추가적 계산에 대해 증분되고, 그리고 충족 변수는 다음의 N DDN 사이클 위상의 사이클 동안 충족 변수의 상황을 표시한다. 예컨대, 충족 변수는 충족 변수 임계치가 만족 또는 초과되었는지를 표시할 수 있다. 이들 데이터 객체는 DDN이 각각의 반복에서 생성된 DDN-생성된 일시적 데이터 객체와 충족 변수 임계치의 비교를 통해 상태 기반을 보유할 수 있게 한다.

[0079] 고유 식별자인 포인터는 DDN이 그 시퀀스 동안 액세스할 작업 공간 내 하나 이상의 노드를 가리키는데, 아래에서 더 설명될 것이다. 포인터는, DDN의 연산에 또한 중요한, 원시 리스트모드 데이터 및 메타데이터를 포함하고 있는 파일의 위치를 가리킨다.

[0080] DDN에 중요한 메타데이터는 2개의 다른 유형의 참조 노트로부터 유래한다. 우선, 메타데이터는 데이터의 특정 서브세트를 얻기 위해, 일반적으로는 전문가에 의해 정의된 게이트의 형태인, 전문가에 의해 이루어진 결정으로부터 유래할 수 있다. 데이터 서브세트는 계층적 게이트로부터 유래할 수 있다. 구체적 일례에서, 선행하는 게이트의 XML 계층은 DDN 데이터 구조에 의한 사용을 위해 메타데이터에 표현된 컨텍스트 정보를 제공한다. 전문가 결정에 대한 대안으로, 메타데이터는, 생물학적으로 유의미한, 염료명("CD3-FITC")에 대한 파라미터를 포함하는 부모 FCS 파일로부터의 키워드 메타데이터를 포함할 수 있다. 메타데이터는 원시 데이터와 연관되고, 그리고 원시 데이터와 연관된 메타데이터는 또한 노드 이름 및 분석될 원시 데이터의 소스인 FCS 파일의 헤더를 포함할 수 있다.

[0081] 마지막으로, 원시 리스트모드 데이터는 이벤트/세포당 수집된 n개의 파라미터에 대한 원시 이벤트/세포 레벨 데이터를 포함한다.

[0082] 도 6에서 예시된 방법은 위에서 설명된 모든 데이터 객체를 사용한다. DDN 방법/라이프 사이클은 제1 DDN 사이클 위상으로 시작한다. 제1 DDN 사이클 위상에서, 프로세서는 메모리에 연산 변수를 로딩한다. 후속하여, 프로세서는 위에서 설명된 메타데이터를 로딩한다. 프로세서는 프로세서가 DDN 흐름을 통해 분석될 파일을 로딩하기 전에, 시험을 위한 규칙 및 변수를 정의하는, 메타데이터 및 연산 변수를 로딩한다.

[0083] 연산 변수 및 메타데이터를 로딩한 후에, 프로세서는 메타데이터가 DDN의 기준을 만족하는지 결정하기 위해 메타데이터 규칙 연산 변수(들)에 대비하여 메타데이터를 시험한다. 예컨대, 메타데이터 규칙 연산 변수는, 사용자에 의해 설정된 키워드 메타데이터, FCS 파일에 의해 설정된 표현형 메타데이터, 염료 식별 메타데이터, 또는 획득 기기에 의해 발생된 파일 내에 포함된 어느 다른 메타데이터를 통해, CD4 세포가 존재함을 표시하도록 메타데이터 파라미터를 특정한다.

[0084] 연산 값에 대비하여 메타데이터를 시험하는 것은 느슨한 모드, 보통 모드, 및 엄격한 모드와 같은 복수의 모드를 가질 수 있다.

[0085] 느슨한 모드는 메타데이터 요건을 가지고 있지 않을 수 있다. 느슨한 모드에서, DDN은 메타데이터의 값에 무관하게 실행할 것이다. 예컨대, 느슨한 모드에서 DDN은 제공된 리스트모드 원시 데이터에서의 2개의 포인트 간 국소 최소값을 계산하고, 그 후 DDN은 메모리로의 원시 데이터의 로딩을 야기하고, 계산을 인보크하고, 그리고 사용자에게 표현되도록 작업 공간에 통계를 부가함으로써 완료할 것이다.

[0086] 보통 모드에서, 메타데이터 매치의 임계치는 사용자에 의해 설정된다, 예컨대, DDN에 대한 6개의 파라미터 중 3개가 설정되면, 그때 DDN은 그것이 데이터 공간에서 세포 집단을 식별하기에 충분한 파라미터를 가지므로 실행할 것이다.

[0087] 그리고, 엄격한 모드에서, 모든 메타데이터 요건은 DDN의 실행이 개시하기 위해 만족되어야 하고 그리고 프로세서는 메모리에 원시 데이터를 로딩하지 않고, DDN 방법은 정지하고, 그리고 추가적 계산은 수행되지 않는다.

[0088] 메타데이터는 메타데이터 규칙 연산 값의 기준을 만족하든지 또는 그것은 연산 값에 의해 설정된 기준을 만족하지 않든지 할 것이다. 메타데이터가 연산 값의 기준을 만족하지 않으면, 프로세서는 메모리에 원시 데이터를 로

딩하지 않고, DDN 방법은 정지하고, 그리고 추가적 계산은 수행되지 않는다. 메타데이터가 연산 값의 기준을 만족하면, 프로세서는 메모리에 원시 데이터를 로딩한다. 메모리에 로딩된 원시 데이터는 원시 획득 데이터, 다른 노드로부터의 데이터, 하나 이상의 게이트로부터의 데이터, 또는 분석 소프트웨어가 액세스가능한 어느 다른 원시 데이터의 형태로 올 수 있다.

[0089] 원시 데이터를 로딩한 후에, 프로세서는 연산 변수에 의해 특정된 알고리즘 또는 계산을 실행한다. 예컨대, 프로세서는 원시 데이터 내에서 하나 이상의 집단을 식별하기 위해 여기에서 설명된 플러그-인 아키텍처 및 프레임워크를 사용하여 외부 알고리즘을 실행할 수 있다. 부가적으로, 프로세서는 위에서 설명된 DDN 일시적 객체를 생성한다. DDN 일시적 객체를 생성하는 것은 프로세서가 반복 변수를 시작 번호로 설정하고 그리고 실행된 계산 또는 알고리즘의 결과에 기반하여 충족 값을 정의하는 것을 수반한다. DDN 일시적 객체를 생성한 후에, 제1 DDN 사이클 위상은 완료하고, 그리고 프로세서는 다음의 n DDN 사이클 위상의 실행을 시작한다.

[0090] 다음의 n DDN 사이클 위상에서, 위상은 DDN 일시적 객체를 로딩하고 그리고 DDN 일시적 객체의 충족 값이 연산 변수에 의해 설정된 충족 임계치 또는 충족 기준을 만족 또는 초과하는지 결정함으로써 시작한다. DDN 일시적 객체를 충족 임계치에 비교하는 것은 프로세서가 반복 변수를 DDN의 충족 변수에 비교하는 것을 포함할 수 있다. 예컨대, 충족 변수가 5회 반복하라고 DDN에 명령하고, 그리고 일시적 객체의 반복 변수가 5보다 작으면, 충족 변수는 만족되지 않을 것이고 그리고 DDN은 다시 반복할 것이다. 다른 일례로서, 프로세서는 DDN 일시적 객체 또는 어느 다른 연산 변수가 다음 계산에 대한 "지시"를 특정하였는지 결정할 수 있다. 예컨대, DDN 일시적 객체에 의해 특정된 지시는 메모리에서의 원시 데이터의 소정 서브세트만이 다음 반복에서 사용되어야 한다고 표시할 수 있다. 다른 일례로서, 충족 값은 카테고리에서 이벤트의 백분율을 정의하는 것에 의해서와 같이 - 정확도를 표시하는 값을 포함할 수 있고, 그리고 프로세서는 정확도 숫자를 충족 기준에 비교할 수 있다. 정확도 숫자의 일례는 산란 게이트의 순도 및 회수를 추정하는 3-색 흐름의 분석을 포함할 수 있다. 여기에서 산란 게이트는 순도와 회수의 최상 조합에 도달될 때까지 재정의될 수 있다. 최적화 루프는 순도 효과 및 회수 효과 값이 90%를 넘을 때까지 모든 샘플에 적용된 게이트를 수축 및 성장시킬 것이다.

[0091] DDN 일시적 객체의 충족 변수가 충족 임계치 또는 충족 기준을 만족 또는 초과하면, 프로세서는 완료 위상을 실행한다.

[0092] DDN 일시적 객체의 충족 변수가 충족 임계치 또는 충족 기준을 만족 또는 초과하지 않으면, 프로세서는 다음 반복에 대해 일시적 객체가 메모리에 로딩된 원시 데이터의 서브세트를 좌우하는지 메모리에 로딩된 원시 데이터의 전 세트를 좌우하는지 결정한다. 연산 변수는 계산 또는 알고리즘을 데이터의 서브세트에 관하여 실행할지 데이터의 전 세트에 관하여 실행할지 표시할 수 있음을, 위로부터, 상기하라. 예컨대, 연산 변수는 데이터 세트가 외부 알고리즘을 사용하여 게이팅되어야 하고, 그리고 다운스트림 수학적 계산은 게이팅된 데이터에 관해서만 수행되어야 한다고 표시할 수 있다. 메타데이터는 어느 계산 또는 알고리즘을 적용할지 결정하기 위해 데이터 또는 원시 리스트모드 데이터의 메타데이터를 분석하라고 프로세서에 명령할 수 있음을 유념해야 한다. 메타데이터는 특정된 계산 또는 알고리즘을 실행하기 전에 프로세서에 의해 실행될 브랜칭(branching) 또는 결정 트리에 대해 호출할 수 있다. 예컨대, 프로세서가 원시 데이터를 그것이 CD4 이벤트를 시사하게 되도록 분석하면, 프로세서는 CD4 집단 식별 알고리즘을 적용할 수 있는 반면, 프로세서가 원시 데이터를 그것이 CD8 이벤트를 시사하게 되도록 분석하면, 프로세서는 CD8 집단 식별 알고리즘을 적용할 수 있다.

[0093] 연산 변수가 데이터의 전 세트를 특정하면, 프로세서는 원시 데이터의 전 세트에 관하여 특정된 계산 또는 알고리즘을 실행하고, 그리고 프로세서는 반복 변수를 증분하고 그리고 데이터의 전 세트에 관한 실행된 계산 또는 알고리즘의 결과에 기반하여 충족 값을 재정의함으로써 일시적 객체를 업데이트한다. 데이터의 전 세트는 이들 위상 동안 여전히 메모리에 있을 수 있다. 일시적 객체를 업데이트한 후에, 프로세서는 새로운 일시적 객체 값에 기반하여 다음의 n DDN 사이클을 반복한다.

[0094] 연산 변수가 데이터의 서브세트를 특정하면, 프로세서는 원시 데이터의 특정된 서브세트에 관하여 특정된 계산 또는 알고리즘을 실행하고, 그리고 프로세서는 반복 변수를 증분하고 그리고 데이터의 서브세트에 관한 실행된 계산 또는 알고리즘의 결과에 기반하여 충족 값을 재정의함으로써 일시적 객체를 업데이트한다. 데이터의 특정된 서브세트 내에 포함되지 않은 데이터는 메모리로부터 해제되고 다른 곳에 저장될 수 있다. 일시적 객체를 업데이트한 후에, 프로세서는 새로운 일시적 객체 값에 기반하여 다음의 n DDN 사이클을 반복한다.

[0095] 다음의 n DDN 사이클 위상은 충족 임계치 또는 기준이 만족 또는 초과될 때까지 계속된다. 만족 또는 초과되고 나면, 프로세서는 연산 변수에 의해 특정되는 출력 유형을 프로세서가 결정하는 완료 위상으로 계속된다. 반복 옵션에서, 사용자는 DDN 실행 파라미터로서 저장되는 반복 수를 설정할 수 있다. 이러한 결정에 기반하여, 프로

세서는 작업 공간에서 액션을 취하고 DDN 흐름의 결과를 작업 공간 파일에 기록한다. 예컨대, 프로세서는 결과 및 취해진 액션에 종속하여 복수의 시각화 중 하나를 제시할 수 있거나, 또는 프로세서는 작업 공간 파일 내에서 새로운 집단 또는 통계를 정의할 수 있다.

[0096] 완료 위상에서 일어나는, 작업 공간에서 취해진 액션은 새로운 입력 데이터로 DDN을 다시 인보크하는 것을 수반할 수 있다. 예컨대, 완료 위상 동안 발생된 출력은 새로운 입력 데이터 세트일 수 있다. 입력 데이터 세트가 변경될 때, DDN는 다시 인보크하고 프로세싱을 수행할 수 있다. 그리하여, 입력 데이터 세트가 변경될 때마다, DDN는 그 필요한 프로세싱을 수행할 수 있다.

[0097] 이제, 도 8을 참조하면, 어느 실험, 임상 시험, 연구, 연구 프로젝트 등에서라도, 전문가의 수는 한정되어 있다. 즉, 누군가 연구, 토픽, 세포 표현형, 과학적 속성 등의 영역에 대해 더 많이 알고 있을수록, 그들 전문가는 더 적게 존재하고, 전문가의 시간은 한정되어 있다. 그렇지만, 고도로 숙련되고 정통할 수 있는 그러나 전문가가 소지한 풍부한 지식은 부족한 분석가는 훨씬 더 흔하고 풍부하다. 전문가의 부족 및 분석가의 풍부에 기인하여, 전문가는 일반적으로 실험을 실행하는 것과 같은 일부 태스크를 분석가에게 위임하는 한편, 전문가는 분석가 작업 산물을 감독한다. 그렇지만, 관용적 방법은, 기하학적 게이트가 어떻게 적용되었는지와 같은, 실험 및 분석의 각각의 개개의 단계를 전문가가 볼 수 있게 하지 않았는데 단순히 전문가는 그가 검토하는 모든 실험 분석으로부터의 모든 분석 단계를 검토할 시간이 부족하기 때문이다.

[0098] 전문가 이용의 관용적 방법과는 대조적으로, 도 8은 분석가가 전문적으로 훈련된 분석 흐름을 획득된 데이터 세트에 인보크 및 전개할 수 있도록 전문가에 의해 DDN을 훈련시키는 프로세스를 예시한다. 위에서 언급된 바와 같이, 전문가는 DDN의 연산 데이터 구조를 설정함으로써 그리고 저장된 일시적 객체를 통해 DDN에 의해 얻은 지식을 사용함으로써 DDN 데이터 구조에 훈련을 제공할 수 있고, 그리고, 계층적 게이팅의 형태로와 같은, 전문가의 결정은 DDN의 메타데이터에 저장 및 표현될 수 있다. 도 8은 전문가가 그 자신의 전문성 및 경험을 사용하여 데이터 발견 노드를 훈련시키는 것을 예시한다. 훈련 프로세스는 도 6에서 예시된 단계 중 일부 또는 전부를 포함할 수 있다. 전문적으로 훈련된 DDN은 분석 흐름의 일부 또는 분석 흐름 전체를 표현할 수 있다. 예컨대, 전문적으로 훈련된 DDN은 전문가의 지식에 기반하여 정밀한 기하학적 게이팅 기술을 적용할 수 있다. 대안으로, DDN은 집단 식별을 위한 외부 발견 알고리즘을 호출해 내는 분석 단계를 포함할 수 있고, 그리고 전문적으로 훈련된 DDN은 전문가에 의해 제공된 발견 프로세스에 대한 특정 파라미터를 제공할 수 있다. 전문가가 DDN을 훈련시켰고, 흐름 단계를 특정했고, 집단 식별 상의 제한을 특정했고, 그리고 어느 수학적 모델이라도 특정했기 때문에, DDN은 분석가가 분석에 가져올 수 있는 어떠한 편견도 제거한다. 여기에서 논의된 DDN 프레임워크 및 프로세스로, 분석가 편견은 제거되고, 그리고 DDN을 사용하여 수행된 모든 실험은 동일한 방식으로 수행될 것이며, 그것은 균일한 결과를 가져온다.

[0099] 도 8b는 전문가가 어떻게 DDN을 훈련시킬 수 있는지에 관한 일례를 도시한다. 이러한 예에서, 전문가는 더 넓은 CD4 게이트가 더 양호한 분석 결과를 산출함을 알아차릴 수 있다. 그때, 전문가는 컴퓨터 상의 사용자 인터페이스를 사용하여 그의 DDN에서의 CD4 게이트 정의를 넓힐 수 있으며, 그것은 전문가에서 CD4 집단을 조사하고, 그리고 분석에서 더 많은 CD4+ 세포를 포함하도록 범위 게이트를 편집함으로써 수행된다. DDN을 조절한 후에, 조절된 DDN은 데이터베이스에 저장되게 된다. 분석가는 DDN이 다른 CD4 게이트 정의를 가짐을 알고 있지 않고 조절된 DDN을 인보크할 수 있다. 조절된 DDN을 인보크함으로써, 조절된 DDN에 의해 정의된 분석 흐름 전체는 분석 소프트웨어의 단일 세션에서 일어날 것이다. 조절된 DDN은 조절된 방법에 따라 결과를 발생시킬 수 있다. 이러한 방법의 많은 혜택 중에서, 실질적 혜택은 분석가가 다른 액션을 수행하지 않았더라도 조절된 분석 방법이 전문가에 의해 완전히 유효성 검증됨을 알고 있다는 것이다.

[0100] 다른 혜택으로서, DDN은 그룹 또는 개인들 간 공유될 수 있다. T-세포에서의 전문가는 NK 세포 상에 분석을 실행하기 위해 NK 세포에서의 전문가에 의해 생성 및 최적화된 DDN을 검색할 수 있다. 그리하여, 전문성은 전문가들 간 공유될 수 있고, 그리고 실험은 수많은 표현형 상에 효율적으로 실행될 수 있다.

[0101] 능동 노드의 주 혜택 중 하나는 노드가 특정 데이터 세트로부터 분리되고 데이터-구동식이라는 것이다. 데이터가 분석 흐름을 구동하기 때문에, 이용가능하게 되는 분석의 유형은 입력 데이터의 선택에 종속하여 다를 것이다. 환언하면, DDN이 무엇을 계산 및 실행할 수 있는지는 입력 데이터에 종속한다. 일반적으로, 입력 데이터는 과학적 데이터를 표현하는 이벤트 세트, 또는 등가의 내포적 의미를 갖는 파일 세트이다. 예컨대, 입력 데이터는 다수 시점에 걸쳐 캡처링된 CD3 측정치일 수 있다. 다른 일례로서, 입력 데이터는 획득 기기에 의해 캡처링된 원시 데이터일 수 있다. 다른 일례에서, 입력 데이터는 분석 소프트웨어 또는 외부 알고리즘에 의해 발생된 결과적 데이터일 수 있다.

[0102] DDN의 메타데이터는 또한 제약, 브랜칭, 결정 트리를 적용할지, 자체-최적화할지, 또는 실시간으로 반복할지 특정할 수 있으며, 그것은 사용자에 의해 특정되고 DDN 실행 파라미터로서 저장된다. 입력 분석 단계는 DDN이 분석 단계 후에 피드백되는 결과 데이터를 수신함에 따라 수많은 횟수 일어날 수 있다. 브랜칭할지, 제약을 적용할지, 결정 트리를 적용할지 등은 충족 변수 또는 DDN의 메타데이터 내에서 설정될 수 있다.

[0103] DDN이 제약을 적용할 때, DDN은 데이터의 범위를 좁힌다. 예컨대, 좁힐 입력 데이터가 단일 파라미터 분포이었으면, 제약은, 범위가 1 내지 100인 이벤트와 같은, 범위일 수 있다. 범위를 좁힘으로써, DDN은, 잔해이거나 상당한 잡음을 부가할 수 있는, 극단 통에서의 세포를 배제할 수 있다. DDN의 컨텍스트에서의 제약의 다른 적용은, 환자에서의 T-세포 유형의 비가 "역전되는" HIV T-세포 역전 또는 낮은 백혈구 카운트와 같은, 2개의 표현형의 비 또는 서브세트의 빈도를 계산하기 위해 잡음을 제거하는 것일 것이다. 예컨대, 제약은 데이터의 제약된 서브세트에만 계산을 수행하도록 연산 변수를 설정함으로써 적용될 수 있다.

[0104] DDN이 브랜칭을 적용할 때, DDN은 결과가 후속 실행 단계에 영향을 미칠 작업 흐름에서의 포인트를 발생시킨다. 단순한 일례로서, DDN이 CD3+ 서브세트를 찾으려고 시도하고 있지만, DDN이 CD3+ 이벤트가 없다고 결정하면, 그 정보는 프로세스에서 사용될 수 있고 그리하여 다운스트림 분석을 적응적으로 전향한다. 이러한 예에서, DDN은 CD3+ 세포를 찾으려고 검색하기 위해 집단 식별 알고리즘을 적용할 수 있다. DDN은 CD3+ 세포가 발견되지 않았음을 식별시키는 클러스터 집단 결과를 수신할 수 있다. DDN은, 도 5의 피드백 루프를 표현하는, 집단 식별 알고리즘의 결과를 분석하고, 그리고 CD3+ 세포에 관한 보고서를 발생시키는 단계가 쓸모 없을 것이라고 결정할 수 있다. 그래서, 대신, DDN은 새로운 집단을 식별하도록 집단 식별 알고리즘에 요청할 수 있다. 위에서 논의된 HIV 역전 예에서, DDN이 DDN에 의해 로딩된 메타데이터를 사용하여 HIV 역전 상황을 검출하면, DDN은 T-세포 수의 더 상세한 보고 또는 T-세포 수가 정상적 범위에 있었다는 보고를 수행하도록 분석 소프트웨어에 명령할 수 있다. 브랜칭 문장의 사용은 인-세션 프로세싱을 바꿔, 적응적 실행 및 인-메모리 데이터 양자를 레버리징하는 것을 가능하게 한다. 연산 변수는 완료 위상 동안 브랜칭의 이러한 유형을 특정할 수 있다. 대안으로, 메타데이터는 데이터의 서브세트에든 전 세트에든 적용된 특정된 계산 또는 알고리즘을 변경하는 고유 브랜칭을 포함할 수 있다.

[0105] DDN은 특정 결과를 구하기 위한 프로세싱 흐름 전체의 표현인 결정 트리를 적용할 수 있다. 예컨대, 도 9는 특정 집단을 찾기 위한 결정 트리 예를 예시하는데, 이벤트 또는 차원 축소를 수반할 것이다. 도 9에 관한 소정 컨텍스트에 대해, 샘플 데이터 세트에서 특정 유형의 이상을 검출하기 위한 여러 전문화된 패널이 이미 존재한다. 이들 패널이 표현하는 특정 표현형의 조합은 마커를 포함하고 있는 데이터로부터 마이닝될 수 있다. DDN은 이러한 로직에 대한 컨테이너로서 역할을 할 수 있다. 이러한 배열은 패널이 대부분 불필요하게 될 포인트에 도달하게 되도록 패널-기반 분석으로부터 알고리즘 분석으로의 점프를 가능하게 한다. 다수 패널을 시험하기보다는, 관련 있는 마커 전부를 하나의 튜브에 포함시킬 수 있는데, DDN의 정교한 프로세싱 능력이 그러한 시험으로부터 생기는 다수의 데이터 파라미터를 통해 내비게이팅하도록 사용될 수 있음을 의미한다.

[0106] 도 9를 참조하면, 입력 데이터는 파일 모음 또는 게이트일 수 있다. 도 9로부터 알 수 있는 바와 같이, DDN은 CD45+ SSC 데이터가 이용가능한지 결정할 수 있다. 예이면, DDN은 현저한 "어둑한" CD45 피크가 있는지 결정하기 위해 데이터를 분석한다. 순차로, DDN은 다음의 비교 및 분석을 실행한다:

[0107] a. 도 9에서의 2개의 브랜치는, 분석이 완료될 수 있는지 우선 조사하기 위해, DDN이 수행하는 유효성 검증의 프로세스를 예시한다(DDN 실행 파라미터에 비교 - 이 경우에서는, 샘플이 SSC 및 CD45 파라미터를 포함하고 있는가?

[0108] b. 그렇다면, 그때, 지식 베이스로부터의 전문가 게이트는 SSC 및 CD45 파라미터에 의해 식별된 집단에 적용된다.

[0109] c. 피크 찾기(집단 식별) 알고리즘이 실행되어 CD45 파라미터만을 조사하여 (이미 게이팅된 CD45+ 집단에 비해) CD45 어둑한 피크가 있는지 본다.

[0110] i. 피크가 존재하면, 그때는 다른 전문가 계열의 계층적 게이트가 적용되어, 이 경우에서는 급성 단핵구 백혈병(AML) 모구를 식별한다.

[0111] d. 무관하게, 림프구 상에 CD19+ 빈도가 이상하게 높은지(DDN 생성시 전문가에 의해 정의된 바와 같은, 2개의 표준 편차보다 더 큰지) 조사하기 위해 CD19+ 세포는 지식 베이스에 비교되어 DDN-적용된 집단 정의에 의해 식별된다. 림프구 상에 CD19 빈도가 이상하게 높으면, DDN은 카파/람다, CD10, CD22 등처럼 B-세포 특정 정보에 대한 분석을 적용할 것이다.

[0112] 모든 경우에, DDN은 유효성 검증을 수행하고(분석이 수행될 수 있음), 분석의 위상을 실행하고(예컨대, 이들 전문가-정의된 기하학적 게이트를 적용하거나 피크 찾기를 수행하고), 생물학적 결과에 비교하고, 그리고 반복할 수 있다. 이러한 방식으로, DDN은 분석을 지시하도록 그 3개의 정보 유형을 레버리징한다.

[0113] 도 9에서의 비-한정적 예로부터 알 수 있는 바와 같이, DDN은 결정 트리에서의 각각의 포인트에서의 결정의 결과에 기반하여 프로세싱을 변경할 수 있다. 프로세싱에서의 변경은, 예컨대, 유효성 검증 기준이 실패할 때, 사용자 인보크시, 연산 변수 및 DDN의 메타데이터에 의해 표현될 수 있다. 또한, 관용적 방법과는 대조적으로, 도 9에서 도시된 결정 트리는 DDN이 분석 소프트웨어의 단일 세션에서 모든 이들 결정 및 결과를 프로세싱하기 때문에 인간에 의한 주관적 편견을 제거한다.

[0114] DDN은 또한 다수의 분석 "패스"에 걸쳐 결과를 정제하도록 최적화 기술을 사용할 수 있다. 최적화의 일례는 세포 분할 사이클의 위상에서 얼마나 많은 세포가 있는지의 추정치를 분석 소프트웨어가 계산하는 세포 사이클 핏팅 분석일 것이다. 분할 사이클에서의 세포의 정확한 수는 계산에서 구한 수를 정제하도록 반복적으로 최상으로 구해진다. 정제 및 최적화는 다수의 패스에 대해 호출되고, 그리고 DDN은 사용자가 정확한 결과를 계산하는데 필요한 "패스"의 수에 관한 제한을 설정할 수 있게 한다. 제한은 반복의 수이거나, 계산 정확도 개선이 개선 임계치를 초과해야 하거나 프로세스가 중단되게 하는 임계치 델타를 사용하는 것일 수 있다. 세포 사이클 핏팅 분석은 집단 식별로까지 확장될 수 있는데 여기서 식별 알고리즘은 식별 기술이 개선 임계치 델타를 더 이상 초과하지 않을 때까지 반복적으로 표현형을 할 수 있다. 프로세서는 최적화 기술에 기반하여 DDN 메타데이터를 변경할 수 있다.

[0115] 더욱, DDN은 각각의 단계 후 차원 또는 파라미터 범위를 축소하면서 프로세스를 반복하도록 반복을 사용할 수 있다. 예컨대, DDN은 데이터 범위의 최소값 또는 최대값으로부터 시작하여 분석함으로써 데이터의 분포에서 모든 피크(최대값)를 찾을 수 있다. 제1 피크를 찾고 나면, DDN은 데이터 세트로부터 피크를 제거하여 DDN이, 최대값으로부터의-제2 피크 등과 같은, 더 많은 피크를 찾을 수 있게 한다. 결국, 하나의 피크만이 남을 것이고, 그리고 DDN이 마지막 피크를 찾은 후에, 반복은 정지한다. 반복은 충족 변수 내에 포함된 반복 변수에 의해 정의될 수 있다.

[0116] 마지막으로, DDN은 다른 유사한 DDN으로부터 학습된 지식 및 훈련을 레버리징할 수 있다. DDN이 전문가에 의해 생성될 때, DDN은 유사한 DDN 데이터 구조에 대해 데이터베이스에 쿼리를 행하도록 구성된다. DDN은 그 메타데이터에서 유사한 이름 또는 유사한 항목을 찾으려고 검색함으로써 이러한 쿼리를 수행할 수 있다. 예컨대, DDN이 그것을 CD4 식별 노드로서 식별시키는 메타-정보를 가지고 있으면, DDN은 유사한 또는 똑같은 메타데이터를 갖는, DDN 데이터베이스에 저장된, 다른 DDN을 찾으려고 검색할 수 있다. DDN은 어느 시맨틱 방법이라도 통해 유사한 DDN을 찾을 수 있다. 유사한 DDN을 찾을시, 새로 훈련된 DDN은 DDN이 이전에 생성된 DDN에 의해 얻은 훈련 및 지식을 수신할 수 있게 할 데이터베이스에 저장된 유사한 DDN으로부터 정보를 얻을 수 있다. 예컨대, 새로 생성된 DDN은 유사한 DDN이 전문적으로 정의된 기하학적 게이트, 또는 게이트의 최소/최대 범위, 게이트에 대한 백분위수, 또는 임상적으로 유의미한 결과를 발생시키는데 도움을 주는 수학적 관계를 가지고 있음을 알아낼 수 있다. 각각의 DDN은 그것이 데이터에 적용되었던 횟수를 다른 DDN 데이터 구조에 통신할 수 있다. 위에서 언급된 바와 같이, 획득된 데이터에 DDN이 더 많이 적용될수록, DDN이 발생시키는 결과는 더 양호하다. 그래서, 더 많은 데이터에 적용된 DDN은 유사한 데이터 구조가 "더 오래된" DDN의 훈련을 레버리징할 수 있도록 다른, 유사한 DDN 데이터 구조에 범위, 백분위수, 게이트, 수학적 관계, 파라미터 프루닝, 또는 어느 다른 중요한 지식이라도 통신할 수 있다. DDN은 인보크를 통해 그리고 또한 데이터베이스에서의 다른 유사한 DDN 데이터 구조와의 통신을 통해 학습하고, 그리하여 최적의, 예컨대, 집단 식별을 내놓도록 반복적 실험 및 전문가의 네트워크를 레버리징한다. 또 다른 일례에서, DDN은 획득 기기에 의해 데이터가 수집되는 방식을 변경하거나 그 방식에 대한 변경을 시사할 수 있다.

[0117] DDN은 컴퓨터의 메모리에서 그리고 메모리에 저장된 입력 데이터 상에서 동작한다. 사용자가 DDN을 사용하려는 제스처를 취할 때, DDN은 필요한 입력 데이터를 메모리에 모으고 그리고 메모리 내 입력 데이터 상에 데이터 프로세싱을 수행한다. 데이터는 DDN이 반복하거나, 제약을 적용하거나, 결정을 하거나, 브랜칭하거나, 최적화함에 따라 축소 및 프루닝될 수 있다. DDN이 더 많은 지능을 얻을수록, DDN은 메모리에 저장된 데이터의 양이 최소화되도록 입력 데이터 상에 초기 프리-프로세싱을 수행할 수 있다. 메타-데이터 레벨에서 일어나는, 데이터 프리-프로세싱에 의해, 컴퓨터의 성능은 DDN이 훈련시키기를 계속함에 따라 증가한다. 더욱, 수동, 기하학적 게이팅의 주관적으로 편견 있는 단계를 제거함으로써, 결과는 이전 실험 방법보다 더 빨리 사용자에게 제시된다. 획득 컴퓨터, 분석 컴퓨터, 또는 서버는 DDN의 모든 특징을 수행하도록 부가적 프로세싱을 수행할 수 있지만, 효율은

DDN의 사용으로 증가된다.

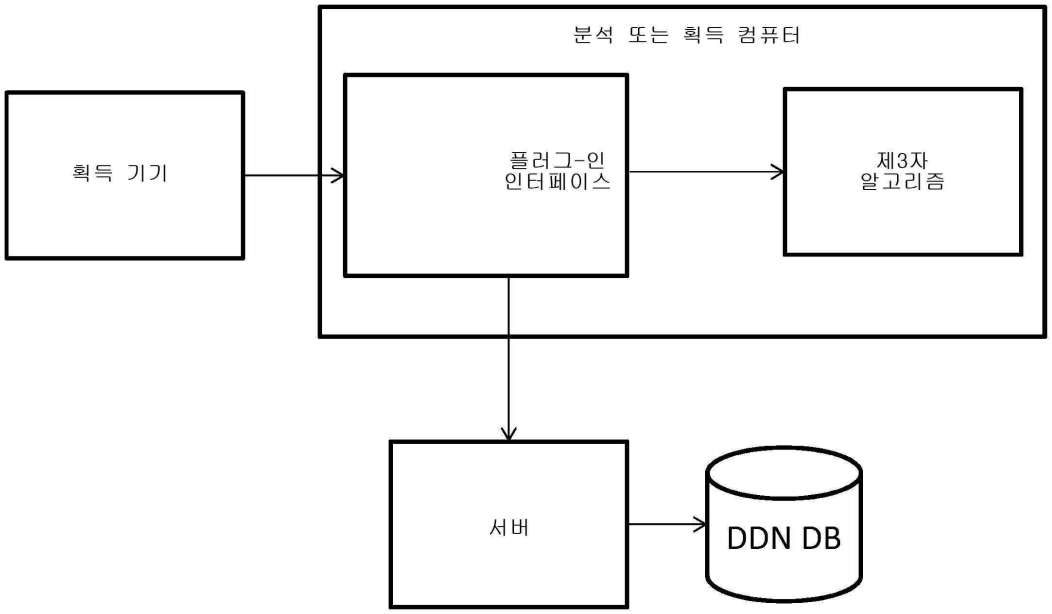
[0118] DDN은 또한 사용자에게 결과를 제시하기 위해 분석 소프트웨어 내에 포함되어 있는 레이아웃 편집기 또는 표 편집기를 레버리징할 수 있다. 일부 컨텍스트에서, DDN은 분석가와 같은 사용자가 단순히 DDN을 인보크하고 어떠한 다른 단계도 없이 분석 소프트웨어를 통해 실험 결과를 제시받을 수 있게 되도록 분석 흐름 전체를 캡슐화할 수 있다. 이러한 식으로, DDN은 실험 전체를 포함하고 있을 수 있다.

[0119] 전술한 것을 고려하여, 본 발명의 수 개의 이점이 달성 및 획득됨을 알 것이다.

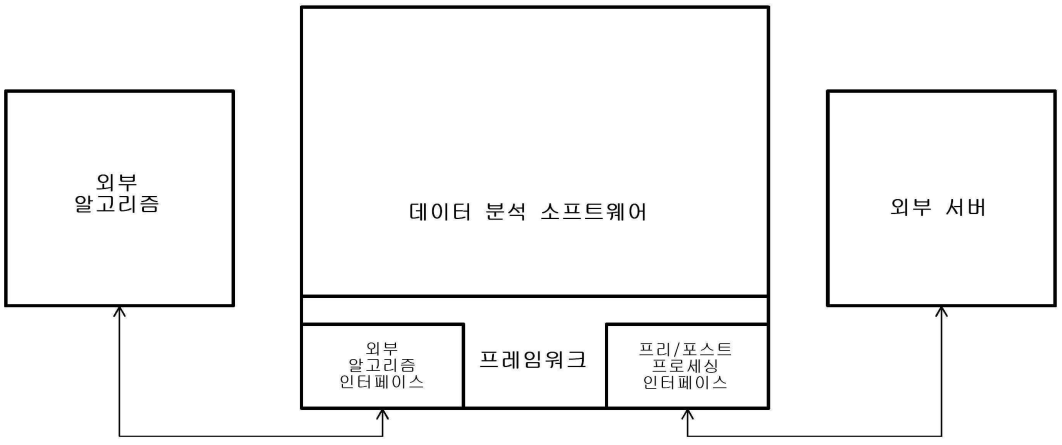
[0120] 본 발명의 원리 및 그 실제 응용을 최상으로 설명하여 그로써 당업자가 본 발명을 다양한 실시형태에서 그리고 고려되는 특정 사용에 어울리는 바와 같은 다양한 수정으로 최상으로 이용할 수 있게 하기 위해 실시형태가 선택 및 설명되었다. 본 발명의 범위로부터 벗어나지 않고 다양한 수정이 여기에서 설명 및 예시된 구성 및 방법에서 이루어질 수 있으므로, 전술한 설명에 포함되거나 수반 도면에서 도시된 모든 사항은 한정적인 것이라기보다는 예시적인 것으로 해석되어야 함을 의도한다.

도면

도면1



도면2a



도면2b

예시적 작업 공간 XML

캡슐화된 데이터 소스, 분석, 결과, 시각화, 및 플러그인 입력

```
<Workspace>
    <Groups>
        ... 분석
    </Groups>
    <SampleList>
        <Sample>
            ... 데이터 소스 참조
            <Population>
                ... 분석 결과
                <Plugin>
                    ... 플러그인 입력
                </Plugin>
            </Population>
        </Sample>
    </SampleList>
    <TableEditor>
        ... 표 보고서 발생
    </TableEditor>
    <LayoutEditor>
        ... 시각화 및 배치 발생 규격
    </LayoutEditor>
</Workspace>
```

도면2c

예시적 플러그인 XML

플러그인은 그들의 작업을 수행하는데 필요한 모든 입력을 저장한다

예. **DownSample**

```
<Sample>
    <SampleNode>
        <Subpopulation>
            <ExternalPopNode>
                <DownSample numEvents="8000" version="1.0"/>
```

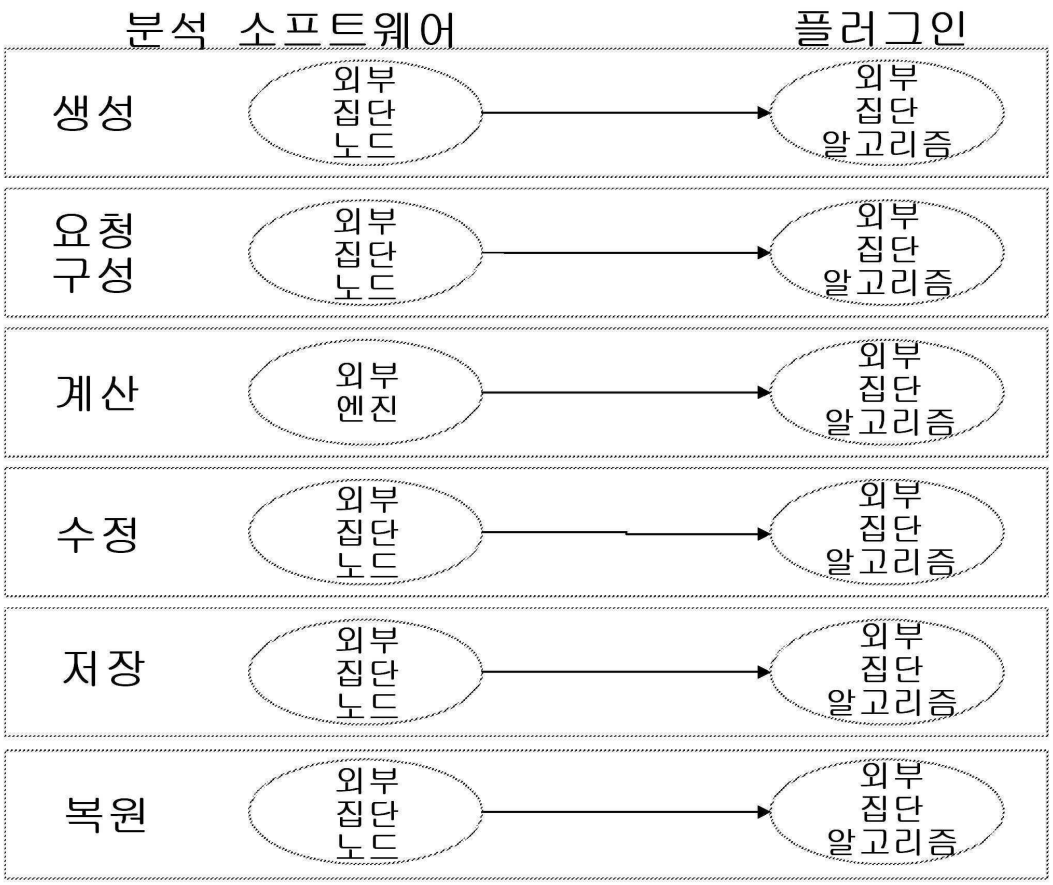
예. **TSNE**

```
<Sample>
    <SampleNode>
        <Subpopulation>
            <ExternalPopNode>
                <TSne version="1.0"/>
                    <Option NumberOfIterations="1000"/>
                    <Option Perplexity="20"/>
                    ...
                    <Parameter name="Rh103di"/>
                    <Parameter name="Xe131Di"/>
                    ...
```

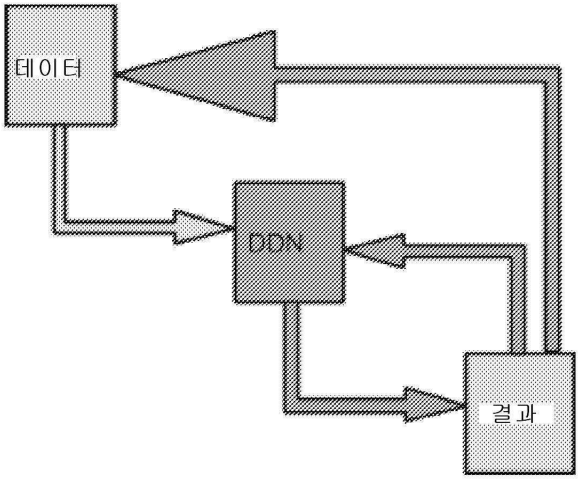
도면3


분석 소프트웨어
플러그인 프레임워크
열기
FlowJo
작업 공간
서버플러그인
저장
FlowJo
작업 공간
서버플러그인
종료
FlowJo
애플리케이션
서버플러그인

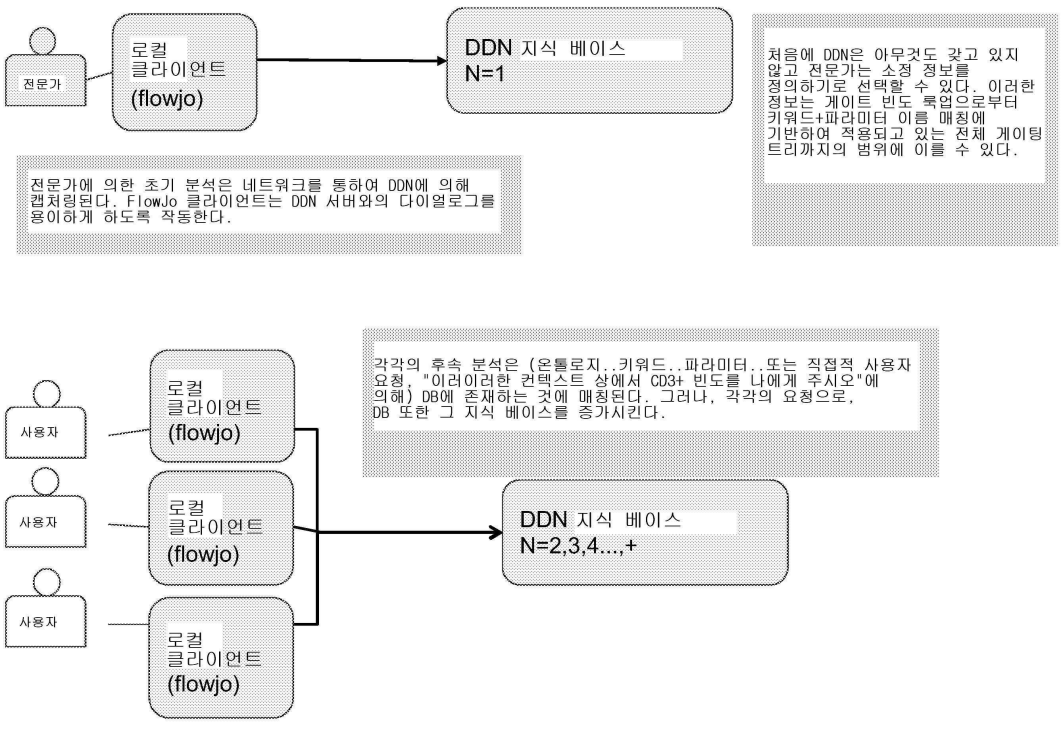
도면4


분석 소프트웨어
플러그인
생성
외부
집단
노드
외부
집단
알고리즘
요청
구성
외부
집단
노드
외부
집단
알고리즘
계산
외부
엔진
외부
집단
알고리즘
수정
외부
집단
노드
외부
집단
알고리즘
저장
외부
집단
노드
외부
집단
알고리즘
복원
외부
집단
노드
외부
집단
알고리즘

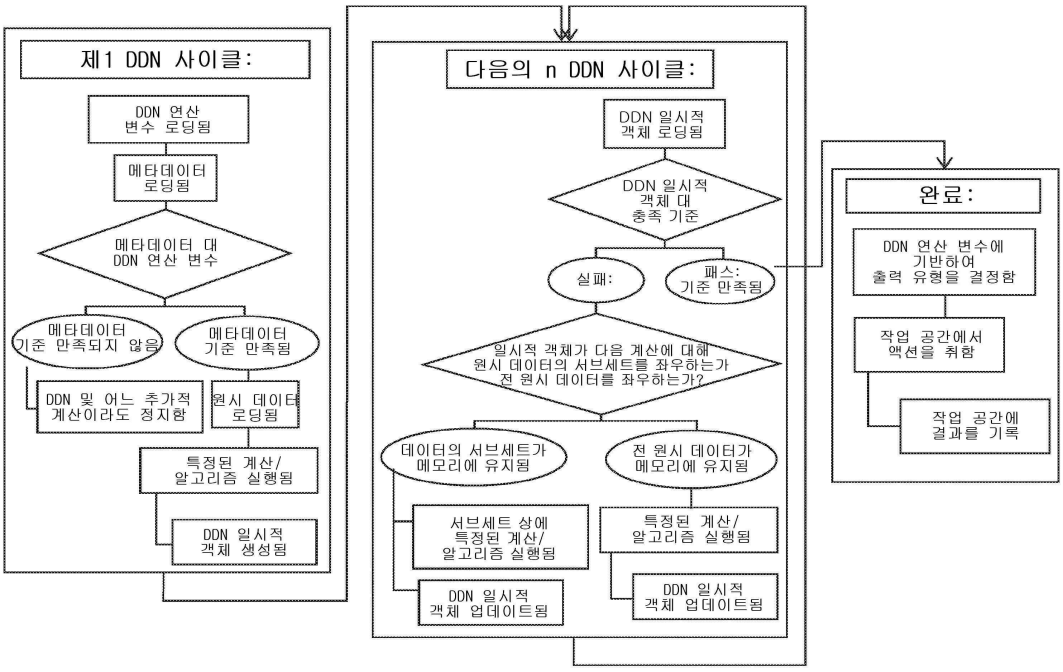
도면5a


데이터
DDN
결과

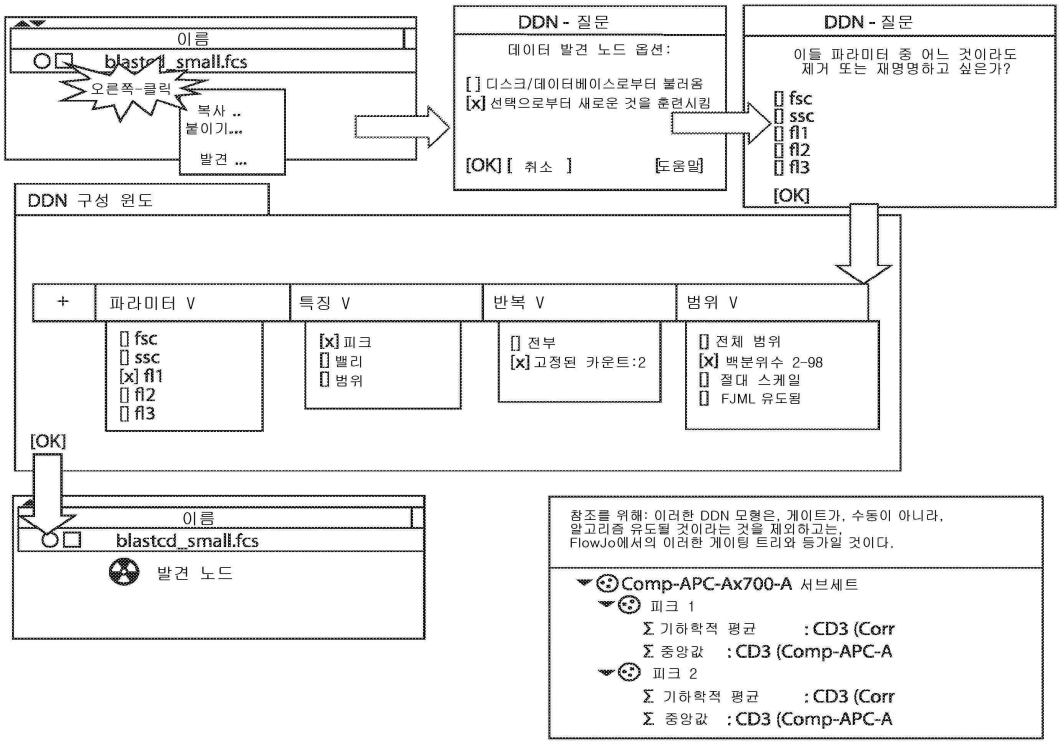
도면5b


전문가
로컬
클라이언트
(flowjo)
DDN 지식 베이스
N=1
처음에 DDN은 아무것도 갖고 있지 않고 전문가는 소정 정보를 정의하기로 선택할 수 있다. 이러한 정보는 게이트 빈도 룩업으로부터 키워드+파라미터 이름 매칭에 기반하여 적용되고 있는 전체 게이팅 트리까지의 범위에 이를 수 있다.
전문가에 의한 초기 분석은 네트워크를 통하여 DDN에 의해 캡처링된다. FlowJo 클라이언트는 DDN 서버와의 다이얼로그를 용이하게 하도록 작동한다.
사용자
사용자
사용자
로컬
클라이언트
(flowjo)
로컬
클라이언트
(flowjo)
로컬
클라이언트
(flowjo)
각각의 후속 분석은 (온톨로지..키워드..파라미터..또는 직접적 사용자 요청, "이러이러한 컨텍스트 상에서 CD3+ 빈도를 나에게 주시오"에 의해) DB에 존재하는 것에 매칭된다. 그러나, 각각의 요청으로, DB 또한 그 지식 베이스를 증가시킨다.
DDN 지식 베이스
N=2,3,4...,+

도면6


제1 DDN 사이클:
DDN 연산 변수 로딩됨
메타데이터 로딩됨
메타데이터 대 DDN 연산 변수
메타데이터 기준 만족되지 않음
메타데이터 기준 만족됨
DDN 및 어느 추가적 계산이라도 정지함
원시 데이터 로딩됨
특정된 계산/ 알고리즘 실행됨
DDN 일시적 객체 생성됨
다음의 n DDN 사이클:
DDN 일시적 객체 로딩됨
DDN 일시적 객체 대 충족 기준
실패:
패스: 기준 만족됨
일시적 객체가 다음 계산에 대해 원시 데이터의 서브세트를 좌우하는가 전 원시 데이터를 좌우하는가?
데이터의 서브세트가 메모리에 유지됨
전 원시 데이터가 메모리에 유지됨
서브세트 상에 특정된 계산/ 알고리즘 실행됨
특정된 계산/ 알고리즘 실행됨
DDN 일시적 객체 업데이트됨
DDN 일시적 객체 업데이트됨
완료:
DDN 연산 변수에 기반하여 출력 유형을 결정함
작업 공간에서 액션을 취함
작업 공간에 결과를 기록

도면7

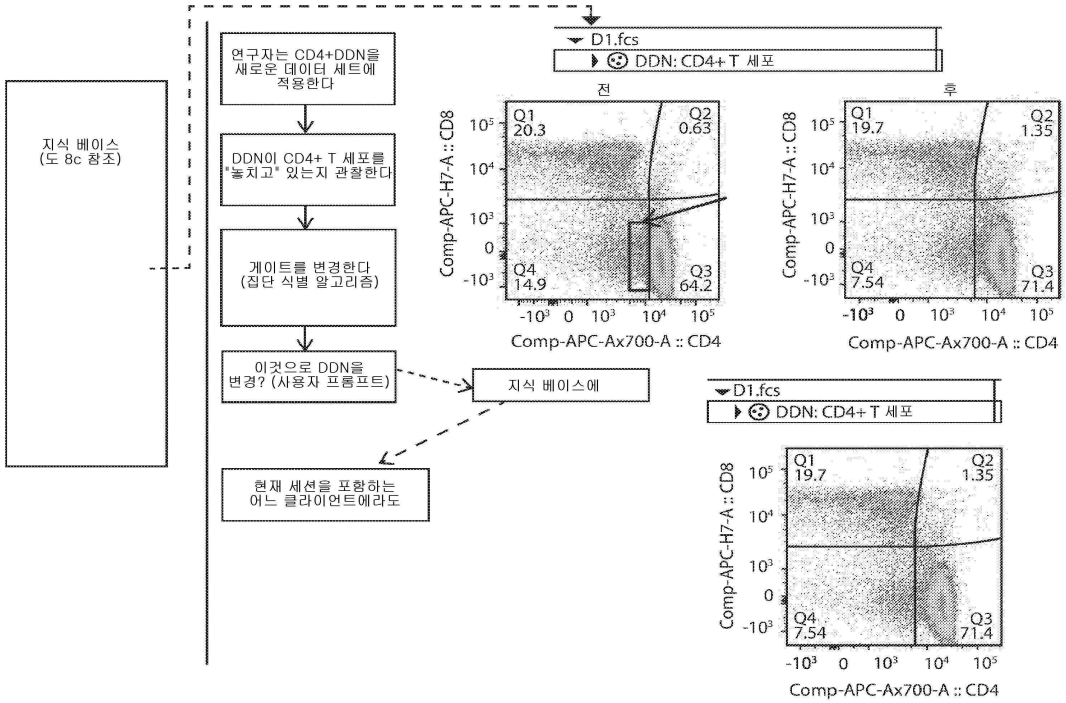

이름
blastA_small.fcs
오른쪽-클릭
복사 ..
붙이기...
발견 ...
DDN - 질문
데이터 발견 노드 옵션:
[ ] 디스크/데이터베이스로부터 불러옴
[x] 선택으로부터 새로운 것을 훈련시킴
[OK] [취소] [도움말]
DDN - 질문
이들 파라미터 중 어느 것이라도 제거 또는 재명명하고 싶은가?
[] fsc
[] ssc
[] fl1
[] fl2
[] fl3
[OK]
DDN 구성 윈도
+
파라미터 V
특징 V
반복 V
범위 V
[] fsc
[] ssc
[x] fl1
[] fl2
[] fl3
[x] 피크
[] 밸리
[] 범위
[] 전부
[x] 고정된 카운트:2
[] 전체 범위
[x] 백분위수 2-98
[] 절대 스케일
[] FJML 유도됨
[OK]
이름
blastcd_small.fcs
발견 노드
참조를 위해: 이러한 DDN 모형은, 게이트가, 수동이 아니라, 알고리즘 유도될 것이라는 것을 제외하고는, FlowJo에서의 이러한 게이팅 트리와 등가일 것이다.
Comp-APC-Ax700-A 서브세트
피크 1
Σ 기하학적 평균 : CD3 (Corr
Σ 중앙값 : CD3 (Comp-APC-A
피크 2
Σ 기하학적 평균 : CD3 (Corr
Σ 중앙값 : CD3 (Comp-APC-A

도면8a

DDN 작업 흐름: 전문가는 DDN을 훈련시키고, 그래서 어느 수의 분석가라도 그것을 이용할 수 있다. 데이터 분석은 편견이 없고 일관되고, 분석가에 의해 도입되는 오류가 없다. 전문가는 일관된, 등가의 결과로의 액세스를 가져서 결정을 하든지, 그 정보를 사용하여 DDN을 훈련시키든지 한다.



도면8b



도면8c

지식 베이스가 포함하는 것:

1. 샘플 정보

샘플 메타데이터

날짜 : 10-OCT-2008
시스템 :Windows XP 5.1
세포측정기 :LSRII
파일:B1 .fcs
파일: URI:file:/Users/mikestadnisky/Deskto
~~~~~~~~~~~~~~~
$BEGINDATA: 3584
$BEGINSTEXT: 0
$BTIM: 11:52:08
$BYTEORD: 4,3,2,1
$CYT: LSRII
$DATATYPE: F
$DATE: 10-OCT-2008
$ENDDATA: 2231503
$ENDSTEXT: 0
$ETIM: 11:52:32
$FIL: B1 .fcs

포함하는 파라미터:

FSC-A
FSC-H
FSC-W
SSC-A
SSC-H
SSC-W
시간
Comp-FITC_Alexa 488-A :: CD56
Comp-PE-A :: CD11c
Comp-PE-Cy55-A :: CD3
Comp-PE-Cy7-A :: CD45RO
Comp-CD3 CD14 Live_Dead-A :: 7AAD
Comp-APC_Ax 647-A :: 2H2
Comp-APC-Ax700-A :: CD4
Comp-APC-H7-A :: CD8

2. DDN 실행 파라미터

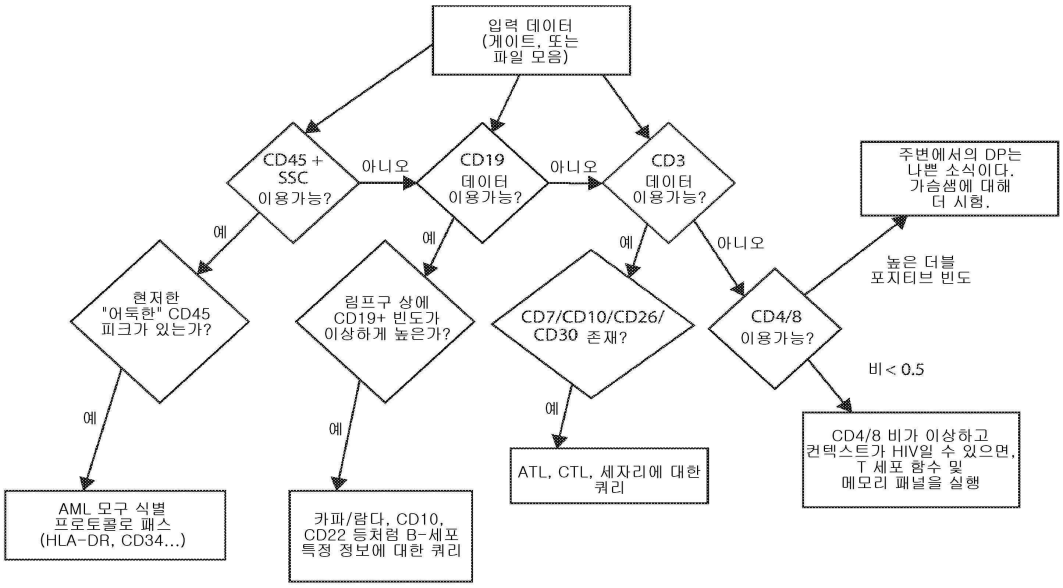
분석 계층(순서), 분석 알고리즘, 파라미터

이름	통계	#세포	그래프 규격	게이트 좌표	게이팅 파라미터	변환 설정
B1.fcs		37132	의사 색 FSC-A() vs. FSC-W(null): tnfTNL			< 변환 >...
단일체	88.7	32950	히스토그램* Comp-PE-Cy55-A(null): tnfTNL	{36122, 87710} {251...	FSC-A 대 FSC-W	
CD3+T 세포	71.3	23479	의사 색 Comp-APC-Ax700-A() 대 Comp-APC	{1448, 0.55} {97848,...	Comp-PE-Cy55-A	
Q1: CD4-, CD8+	17.6	4125	의사 색 Comp-APC-Ax700-A() 대 Comp-APC	2{-, -} {3814, 2293}	Comp-APC-Ax700-A...	
Q2: CD4+, CD8+	1.75	412	의사 색 Comp-APC-Ax700-A() 대 Comp-APC	3{3814, -} {-, 2293}	Comp-APC-Ax700-A...	
Q3: CD4+, CD8-	74.8	17555	의사 색 Comp-APC-Ax700-A() 대 Comp-APC	1{3814, 2293} {-, -}	Comp-APC-Ax700-A...	
Q4: CD4-, CD8-	5.91	1387	의사 색 Comp-APC-Ax700-A() 대 Comp-APC	0{-, 2293} {3814, -}	Comp-APC-Ax700-A...	

3. 생물학적 결과 정보 – 참조 통계

조상 서브세트 통계	부모의 단일체 빈도	부모의 단일체 CD3+ T 세포 빈도	단일체/CD3+T 세포 Q1: 부모의 CD4−, CD8+ 빈도	단일체/CD3+T세포 Q2: 부모의 CD4+, CD8+ 빈도	단일체/CD3+T세포 Q3: 부모의 CD4+, CD8− 빈도	단일체/CD3+ Q4: 부모의 CD4−, CD 빈도
B1.fcs	88.7	71.3	17.6	1.75	74.8	
~~~~~~~~~~~~~~~

도면9


입력 데이터
(게이트, 또는
파일 모음)
CD45 +
SSC
이용가능?
아니오
CD19
데이터
이용가능?
아니오
CD3
데이터
이용가능?
주변에서의 DP는
나쁜 소식이다.
가슴샘에 대해
더 시험.
예
예
예
아니오
높은 더블
포지티브 빈도
현저한
"어둑한" CD45
피크가 있는가?
림프구 상에
CD19+ 빈도가
이상하게 높은가?
CD7/CD10/CD26/
CD30 존재?
CD4/8
이용가능?
비 < 0.5
예
예
예
AML 모구 식별
프로토콜로 패스
(HLA-DR, CD34...)
카파/람다, CD10,
CD22 등처럼 B-세포
특정 정보에 대한 쿼리
ATL, CTL, 세자리에 대한
쿼리
CD4/8 비가 이상하고
컨텍스트가 HIV일 수 있으면,
T 세포 함수 및
메모리 패널을 실행