TEMA 2 AIN

## Agentes inteligentes: Conceptos fundamentales.

**1.1 Definición de agente:**

Los agentes inteligentes se utilizaron en inteligencia artificial principalmente para dos campos:

-Resolución de problemas distribuidos: en los cuales cada agente es un nodo que se encarga de cooperar y compartir una cantidad de tareas que tiene prefijadas. El sistema tiene un comportamiento global o de conjunto.

-Sistemas multiagente: los agentes autónomos trabajan juntos con la finalidad de resolver un problema. Los agentes eligen dinámicamente quienes son los encargados de realizar cada tarea. No hay un sistema de control y la comunicación es asíncrona.

Surge también la programación orientada a agentes y los lenguajes de comunicación de agentes. Incluso hay ingenierías del software orientadas a agentes.

Según la rae un **agente** es algo que obra o tiene virtud de obrar, el que realiza una acción o persona o cosa que produce un efecto, dentro de muchas otras.

Los **agentes software** son aplicaciones informáticas con capacidad para decidir cómo deben actuar ellas mismas para alcanzar sus objetivos.

Los **agentes inteligentes** son agentes software que pueden funcionar fielmente en un entorno cambiable.

La principal característica de un agente es que son autónomos y que son capaces de trabajar independientemente. Una primera definición fue la de un sistema informático que trabaja autónomamente en un entorno para conseguir unos objetivos. Por tanto, tenemos que saber que un agente es una entidad que está constantemente en iteración con el entorno.

Un agente es autónomo, reactivo y proactivo. Tiene una conducta dirigida por el objetivo. Tiene unos actuadores que hacen que el entorno cambie y unos sensores que perciben el entorno y los cambios que ocurren en este.

Según la teoría de agentes, un **agente** es un sistema informático situado en algún entorno, el cual es capaz de percibirlo y, a través de esta percepción, es capaz de determinar (pensar) y ejecutar acciones, de manera autónoma, las cuales le permiten alcanzar sus objetivos. Estas acciones pueden cambiar el entorno.

Un agente inteligente, como bien hemos dicho anteriormente tiene diferentes características, entre ellas:

-Reactividad: el mundo real y, por tanto, el entorno en el que se ejecuta el agente es un entorno cambiante, dinámico. Los programas que se ejecutan en estos entornos deben de ser capaces de reaccionar al fallo y de preguntarse a ellos mismos si les conviene continuar con su ejecución o no. Los **sistemas reactivos** son aquellos que mantienen interacción con su entorno y responden a los cambios que suceden en él.

-Proactividad/iniciativa: reaccionar a un entorno es fácil, lo difícil es la toma de decisiones para llegar al objetivo. Los agentes deben hacer cosas por nosotros, por lo que su comportamiento está dirigido a cumplir el objetivo. La **iniciativa o proactividad** consiste en conseguir tomar los objetivos y que el agente no esté dirigido solamente por eventos (cambios en el entorno), sino que tome la iniciativa.

-Sociabilidad: en algunos casos si los agentes quieren alcanzar sus objetivos deben interactuar con otros. La **sociabilidad** es la capacidad de interactuar con otros agentes (o incluso con humanos) mediante **cooperación, coordinación y negociación.**

**CAPACIDADES SOCIALES DE LOS AGENTES:**

-Cooperación: es trabajar de manera conjunta, como un equipo, para conseguir un objetivo. La mayoría de las veces los agentes no pueden cumplir sus objetivos solos, incluso pueden obtener mejores resultados de manera conjunta.

-Coordinación: consiste en gestionar las interdependencias que hay entre las actividades, por ejemplo, si hay un recurso o finalidad que sea incompatible entre los agentes, necesitamos que estos se coordinen y no haya errores. Es para evitar condiciones de carrera.

-Negociación: capacidad de llegar a acuerdos de interés común. La negociación implica ofertas y contraofertas, con compromisos asumidos por los participantes.

**1.2. Entornos del agente.**

En los entornos complejos, los agentes no tienen el control total del entorno, solo tienen un control parcial, lo que significa que el entorno puede influir en sus acciones. Los entornos no son deterministas, los agentes deben estar preparados para posibles fallos. Los entornos tienen diferentes propiedades:

-Accesible (observable) vs inaccesible (parcialmente observable): observable es aquel en el que el agente puede tener información completa, exacta y actualizada del estado del entorno. Los moderadamente complejos son inaccesibles. Cuanto más accesible un entorno, más fácil diseño de agentes.

-Determinista vs estocástico: un entorno determinista es aquel en el que una acción tiene un solo efecto sin haber incertidumbre a la hora de realizar la acción. Los entornos estocásticos son todo lo contrario.

-Episódico vs secuencial: en los episódicos la actuación de un agente depende de diferentes episodios, sin relación entre el desempeño de un agente en diferentes escenarios. Cada episodio es la percepción del agente y la realización de una única acción posterior. El siguiente episodio no depende de las acciones realizadas anteriormente. En cada episodio, la decisión de acción depende solo del episodio en sí. Los episódicos son más sencillos, porque solo te basas en el episodio actual y no tienes que pensar en los episodios futuros.

-Estático vs dinámico: estático es aquel en el que no se producen cambios, excepto lo que son provocados por las acciones del agente. Dinámico, aquel que tiene procesos que ya operan previamente en él.

-Discreto vs continuo: discreto es en el que hay un número fijo y finito de acciones y percepciones. Continuo todo lo contrario.

**1.3. Agentes como sistemas intencionales.**

Daniel Dennett acuñó el término sistema intencional describiendo las entidades como: “entidades cuyo comportamiento puede ser predicho por el método de atribuir creencias, deseos y perspicacia racional”. Los sistemas intencionales pueden tener varios órdenes. El primer orden son sistemas que tienen creencias y deseos y no tiene creencias y deseos sobre creencias y deseos. Los de segundo orden son más sofisticados y tienen creencias y deseos (u otros estados intencionales) sobre creencias y deseos, tanto sobre el mismo agente como sobre los deseos y creencias de otros agentes. Por tanto, un sistema intencional en informática es aquel en el que los agentes inteligentes tienen aspectos que se consideran del mundo humano, como pueden ser deseos y creencias.

Atribuir actitudes humanas a una máquina es correcto cuando se expresa la misma información sobre la máquina que sobre una persona. Atribuir cualidades mentales es más sencillo para las máquinas con una estructura conocida y sencilla (termostato) pero más complejo y útil para entidades con una estructura que no es completamente conocida.

¿Qué podemos describir con la actitud intencional? Accionar un interruptor es solo la manera que tenemos de mostrar nuestro deseo de encender una luz.

Con sistemas sencillos podemos hacer una explicación mecanicista de su comportamiento, pero esto no lo podemos hacer con los que son más complejos. Necesitamos abstracciones y metáforas para explicar el funcionamiento de los sistemas informáticos actuales. La actitud intencional es una de estas abstracciones.

Los conceptos intencionales son herramientas de abstracción con las que podemos describir, explicar y predecir el comportamiento de los sistemas más complejos. Con esta idea, podemos categorizar a los agentes mediante representaciones anidadas de los mismos (sistemas que incluyen representaciones de otros sistemas)

**-Sistemas post-declarativos:**

En la programación declarativa lo que hacemos es que decimos aquello que queremos conseguir, dando al sistema información sobre las relaciones entre objetos y el mecanismo de control a seguir que se encarga de deducir qué hacer. Con los agentes describimos en alto nivel el objetivo y dejamos que el mecanismo deduzca qué hacer.

**1.4. Arquitecturas abstractas para agentes inteligentes.**

-Agentes y entornos: asumimos que el estado puede estar en un conjunto finito de estados instantáneos discreto (E). Se asume que los agentes tienen un repertorio finito de posibles acciones a realizar que transforman el estado del entorno (Ac)

-Ejecución de un agente:

La ejecución (r) de un agente representa la interacción del agente con su entorno. Es una secuencia de acciones intercaladas.

Una captura de pantalla de un celular

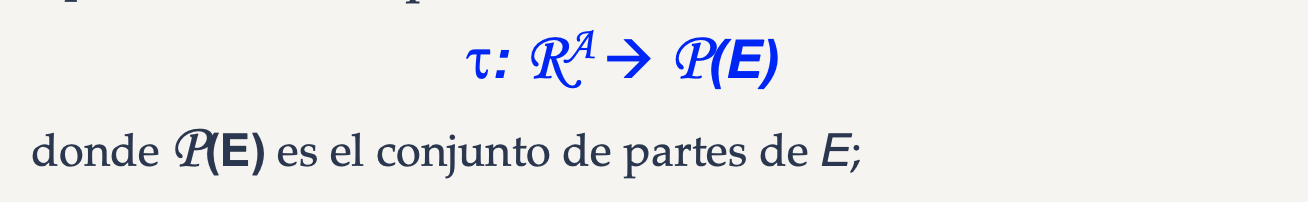
Descripción generada automáticamente

Imagen que contiene captura de pantalla, pájaro, ave

Descripción generada automáticamente

-Entornos:

Estos pueden ser modelados mediante una función de transición de estado que representa el comportamiento del entorno. Esta función tiene como argumento una secuencia de estado-acción que finaliza en una acción y devuelve como resultado un conjunto de estados del entorno, representando el efecto de la acción o acciones del agente. Su significado es que como resultado de ejecutar la ultima de las acciones en un instante dado, el entorno puede encontrarse en una secuencia de estados t(r).



Los entornos son dependientes de la historia y no son deterministas. Si t(r) = 0, no hay posibilidad de alcanzar un estado sucesor de r, entonces la acción ha finalizado. Un entorno Env es una tripleta: Env = <E, s0, t> donde E es el conjunto de estados del entorno, s0 el estado inicial y t la función de transición.