RESUMEN TEORÍA PLANIFICACIÓN Y SISTEMAS COGNITIVOS:

ÍNDICE DE CONTENIDOS:

[] TEMA 1: PLANIFICACIÓN CLÁSICA

} 1.1 ¿QUÉ ES PLANNING?

Planning: Proceso de cálculo de un plan para conseguir uno o varios objetivos.
Plan: Secuencia de acciones que conduce de un estado inicial a un estado que contiene los objetivos a conseguir.
Reglas básicas: El estado del problema está formado por un conjunto de predicados, donde no son válidos ni objetos ni funciones negadas (¬rico y ¬robotAt(bb8,Dagobah)), además de la llamada a una función dentro de otra (areFighting(owner(bb8), DarthVader)) y el uso de objetos sin conexión entre sí (robotAt(x,y)).
Elementos: Acciones/Efectos (permiten cambiar el estado del problema), precondiciones (aplican una acción si se cumplen), Add List (lista de predicados que añaden una acción) y Delete List (lista de predicados que eliminan una acción).

(state, action) = (state - DEL(action)) U ADD(action)

STRIPS (STanford Research Institute Problem Solver): Fue el primer lenguaje de planning ampliamente usado, desarrollado por Imperial College London en 1971 (Shakey en 1984 para el SRI).

} 1.2 PDDL

PDDL 1.2 (Planning Domain Definition Language): Lenguaje oficial del primer IPC (International Planning Competition, 1998) inspirado en STRIPS, cuyo modelo puede aplicarse a una gran variedad de problemas.
Objetivos: Establecer las reglas mediante un dominio, presentar una situación y un goal en el problema, usar un plan solver y conseguir dicho plan.

} 1.3 NUMERIC FLUENTS

Numeric Fluents (PDDL 2.1): Variable que mantiene un valor numérico a lo largo del plan (similar a un predicado).

```
(:functions
  (<variable_name> ?<parameter_name> - <object_type>)
)
```

} 1.4 NUMERIC EXPRESSIONS

Numeric Expressions: (+ (<variable_name>) (<variable_name>)), (/(<variable_name>)(<variable_name>)), (- (<variable_name>) (<variable_name>)), (* (<variable_name>) (<variable_name>)), (increase (<variable_name> ?<parameter_name>) (<variable_name> ?<parameter_name>)), (decrease (<variable_name> ?<parameter_name>) (<variable_name> ?<parameter_name>)) y (assign (<variable_name> ?<parameter_value>) <value>).

} 1.5 DURATIVE ACTIONS

Durative Actions (PDDL 2.1): Simulan cuánto tiempo tardaría el robot en realizar una acción.
Duration (:duration): (= ?duration <duration_number>), (> ?duration <duration_number>), (< ?duration <duration_number>) y (and (> ?duration <duration_number>) (< ?duration <duration_number>)).
Condition (:condition): (at start (<variable_name> ?<parameter_name>)), (at end (>= (<variable_name> ?<parameter_name>) <value>)), (over all (<variable_name> ?<parameter_name>)).
Effect (:effect): (<logical_temporal_condition>).
Continuous Effect: (increase (<variable_name> ?<parameter_name> #t)) y (decrease (<variable_name> ?<parameter_name>) (* 5 #t)).

```
(exists (?<parameter_name> - <object_type>)
        (<variable_name> ?<parameter_name>)
)
```

```
(forall (?<parameter_name> - <object_type>)
        (<variable_name> ?<parameter_name>)
)
```

(when (and <variable_name> ?<parameter_name>) (and <variable_name> ?<parameter_name>))

} 1.6 UPF

UPF (Universal Planning Framework): Proyecto en curso de Horizon Europe que consiste en una librería escrita en Python que proporciona una capa de abstracción para usar múltiples planners.

[] TEMA 2: SISTEMAS DE PLANIFICACIÓN

} 2.1 ¿QUÉ ES UN SISTEMA DE PLANIFICACIÓN?

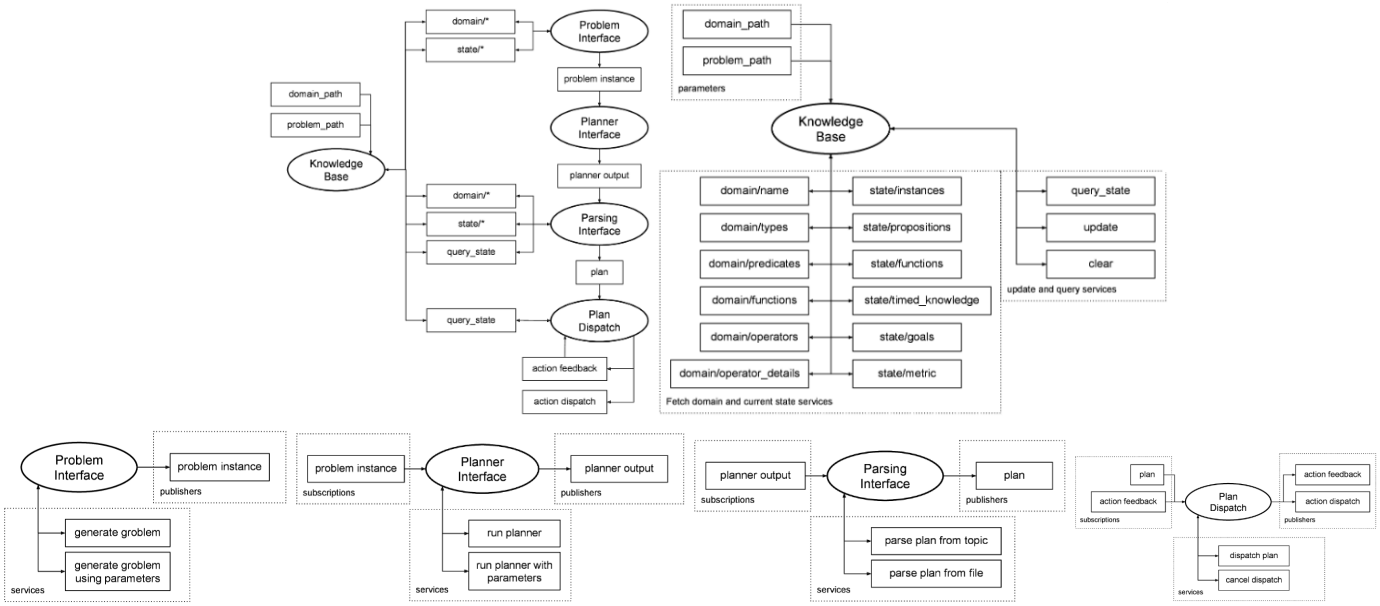
Sistema de Planificación: Sistema que crea y ejecuta planes en robots.
Características: Lee uno o varios dominios PDDL, gestiona el conocimiento de instancias, predicados, goals y funciones, proporciona una interfaz para añadir / eliminar / actualizar el conocimiento, valida con el dominio, proporciona mecanismos para implementar y ejecutar acciones, verifica en tiempo de ejecución los requisitos y aplica los efectos de las acciones.

} 2.2 CORTEX Y SKIROS2

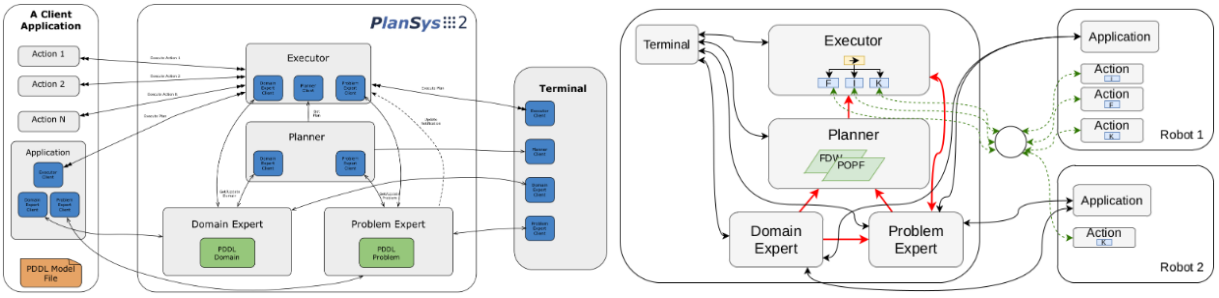
CORTEX: Arquitectura cognitiva que usa Planning, mantiene una representación del conocimiento basada en un grafo donde los agentes acceden al grafo para cumplir sus tareas de percepción, actuación y planning, y un agente de Task Planning genera planes usando Metric-FF.
SKiROS2: Plataforma para crear comportamientos robóticos complejos mediante la composición de skills (bloques de software modulares) en Behavior Trees, donde el usuario proporciona Skills, una escena y un goal, y usa razonamiento basado en ontologías OWL.

} 2.3 ROSPLAN Y PLANSYS2

ROSPlan: Sistema de planificación de referencia en ROS, disponible en ROS.



PlanSys2: Inspirado en ROSPlan, disponible en ROS2, y es eficiente, predecible, seguro y multi-robot.



} 2.4 ARRANQUE DE PLANSYS2

Arranque del sistema: Llevado a cabo por plansys2_bringup en modo monolithic (todos los componentes en un solo proceso, donde las comunicaciones son más rápidas / shared memory, el launcher es menos complejo y un fallo en un componente hace fallar el proceso) o distributed (cada componente está en un proceso y permite arrancar y depurar componentes por separado).
LifeCycle Nodes: Componentes de PlanSys2 orquestados por plansys2_lifecycle_manager, que arrancan primero el domain y el problem expert, por lo que se pueden tener varias instancias de PlanSys2 en diferentes namespaces.

} 2.5 CLIENTES DE PLANSYS2

Clientes: Establecen que cada componente es independiente de ROS2, permitiendo depurar su funcionalidad por separado y futuras migraciones en su estructura. Éstos reproducen una interfaz, donde el propio cliente y el nodo que recubre cada componente ocultan la complejidad de las comunicaciones con servicios.

} 2.6 TERMINAL DE PLANSYS2

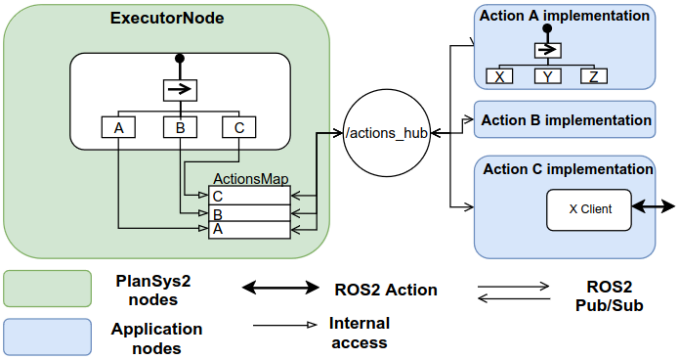
Terminal: Interfaz de shell con funcionalidades avanzadas que permite interactuar con PlanSys2 para gestionar/monitorizar su funcionamiento.

- get
  - model
    - types
    - predicates
    - functions
    - actions
    - predicate [predicate]
    - function [function]
    - action [action]
  - problem
    - instances
    - predicates
    - functions
    - goal
  - domain
  - plan
- set
  - instance [id] [type]
  - predicate [(predicate)]
  - function [(function)]
  - goal [(and(goal))]
- remove
  - instance [id]
  - predicate [(predicate)]
  - function [(function)]
  - goal [(and(goal))]
- run
- check
  - actors

} 2.7 PLANNER Y EXECUTOR DE PLANSYS2

Planner: Pide el dominio y el problema pddl y llama al plan solver, donde cada uno es un plugin (actualmente POPF por defecto y TFD).

Executor: Transforma el plan a un grafo y luego a un BT (es el más complejo de los componentes), cada acción debe tener una implementación a través de un actor (ActionExecutorClient), y las acciones a ejecutar se subastan ya que puede haber más de un actor por acción, actores especializados o multi-robot, los cuales pueden estar implementados por BTs o no.



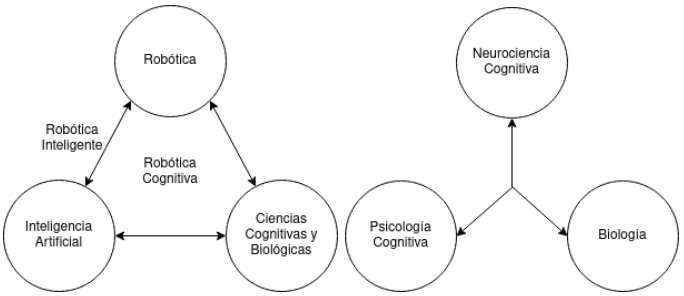
[] TEMA 3: SISTEMAS COGNITIVOS (PARTE 1)

} 3.1 INTRODUCCIÓN

Entornos controlados: El robot se programa para hacer lo que se desee sabiendo lo que nos espera.
Entornos complejos: El robot tiene que ser flexible y adaptable sin saber lo que nos deparará (incertidumbre, conocimiento incompleto y cambios).

} 3.2 ROBÓTICA COGNITIVA

Robótica Cognitiva: Campo que combina conocimientos y métodos de la Robótica, la Inteligencia Artificial y las Ciencias Cognitivas y Biológicas para diseñar un sistema cognitivo integrado que combine el comportamiento sensorimotor, las funciones de alto nivel y las capacidades sociales de un robot inteligente..



Características: Integración a Nivel de Sistema de un rango de habilidades cognitivas (destrezas sensorimotoras, representación de Conocimiento y Razonamiento e Interacción social), Aproximación Interdisciplinaria (neurociencia cognitiva, psicología cognitiva y biología) y bio-inspirado (parecido a un humano y parecido a un animal en cuanto a comportamiento e inteligencia). Se enfoca en diseñar y construir robots que tengan la capacidad de aprender de la experiencia y de otros, memorizar conocimientos y habilidades relevantes, y recuperarlos según lo requiera el contexto para utilizar este conocimiento de manera flexible para seleccionar acciones apropiadas en la búsqueda de sus objetivos, todo ello mientras anticipa el resultado de esas acciones al hacerlo. Por último, su enfoque en la prospección para aumentar la experiencia sensorial-motora inmediata tanto al navegar y manipular objetos en el entorno del robot como al interactuar con personas son una característica clave.

} 3.3 CAMPOS RELACIONADOS

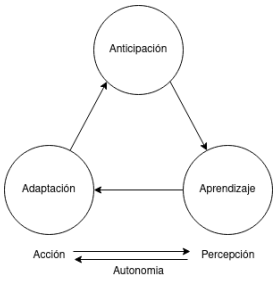
Developmental robotics: Producen un programa que intenta simular la mente de un niño (iCub e Infanoid).
Neurorobotics (Neurorobots): Dispositivos robóticos que tienen sistemas de control basados en principios del sistema nervioso, cuyos modelos operan sobre la premisa que “el cerebro está incorporado y el cuerpo está incrustado en el entorno” (Darwin VII).
Evolutionary robotics (Robótica Evolutiva): Campo de investigación que emplea la computación evolutiva para generar robots que se adaptan a su entorno a través de un proceso análogo a la evolución natural. La generación y optimización de robots se basan en principios evolutivos de variaciones ciegas y supervivencia del más apto, como se materializa en la síntesis neo-Darwiniana.
Soft robotics: Se centra en el diseño, construcción y control de robots flexibles y deformables. Inspirados en la biomecánica de organismos vivos, éstos tienen la capacidad de adaptarse a entornos complejos, interactuar de manera más segura con humanos y objetos delicados, y realizar una variedad de tareas que pueden ser difíciles o imposibles para los robots rígidos tradicionales.

} 3.4 HISTORIA DE LA ROBÓTICA COGNITIVA

Historia de la Robótica Cognitiva: Tortoises (1950), Walter (1953), Shakey (1966), Vehicles y Braitemberg (1986), Darwin (1992), Di Giuseppe (1998), Khepera (1999), CB2 y Pfeifer & Bongard (2007), iCub (2008), y Octopus y Cangelosi & Schlesinger (2015).

} 3.5 COGNICIÓN EN ROBÓTICA

Operaciones autónomas en entornos cotidianos: Los robots se anticipan a los resultados cuando se seleccionan las acciones que llevará a cabo, y se adaptan a los cambios y situaciones imprevistas.
Interacción eficaz con humanos: Una capacidad cognitiva puede permitir que un robot infiera los objetos y las intenciones de la persona con la que está interactuando y por tanto le permite comportarse de manera útil, y a los humanos nos gusta interactuar con otros agentes cognitivos (si un robot tiene la capacidad de cognición, fomenta la interacción del robot humano).
Robot cognitivo: Realiza acciones dirigidas a objetos sensibles al contexto, donde el robot anticipa la necesidad de actuar y el resultado de la acción y la acción en sí está guiada por la prospección y puede también adaptarse a circunstancias cambiantes (ajustar las políticas de acción existentes y crear nuevas políticas de acción cuando sea necesario), todo ello mediante el uso de habilidades cognitivas básicas (percepción, atención, selección de acciones, memoria, aprendizaje, razonamiento, metacognición y prospección).
Cognición: Proceso por el cual un sistema autónomo percibe su entorno, aprende de la experiencia, anticipa el resultado de los eventos, actúa para perseguir metas, y se adapta a circunstancias cambiantes.



Habilidades básicas de un sistema cognitivo: Percepción, atención, selección de acción, memoria, aprendizaje, razonamiento, meta-razonamiento y prospección (capacidad de anticipar el futuro y es, posiblemente, el sello distintivo de la cognición).
Sistemas cognitivos: Predicen contínuamente la necesidad de actuar (uno mismo y los demás) y el resultado de esas acciones, como por ejemplo, las actividades cotidianas (aparentemente rutinarias pero a menudo complejas y exigentes), anticiparse a las necesidades de los demás e interactuar, ayudar y colaborar con otros, donde las acciones están dirigidas a lograr objetivos y se guían por la información prospectiva (futuro).
Robótica cognitiva: Rama de la robótica donde el conocimiento juega un papel central en el apoyo a la selección de acción, ejecución y entendimiento. Se enfoca en diseñar y construir robots que tengan la capacidad de aprender de la experiencia y de otros, memorizar conocimientos y habilidades relevantes, recuperarlos según lo requiera el contexto y utilizar de manera flexible este conocimiento para seleccionar acciones apropiadas en la búsqueda de sus objetivos al tiempo que anticipa el resultado de esas acciones al hacerlo. Los robots cognitivos pueden usar su conocimiento para razonar sobre sus acciones y las acciones de aquellos con quienes interactúan, y por lo tanto, modificar su comportamiento para mejorar su efectividad general a largo plazo. En resumen, los robots cognitivos son capaces de realizar acciones flexibles y sensibles al contexto, sabiendo lo que están haciendo y por qué lo están haciendo.
Objetivos: Percibir su entorno (la percepción hace uso de muchas modalidades sensoriales como visión, audición y háptica), prestar atención a los eventos que importan de forma selectiva (seleccionar una característica u objeto dado), restrictiva (restringir qué buscar o dónde buscarlo) o supresiva (suprimir características, objetos o ubicaciones que se consideran no relevantes), anticiparse a la necesidad de alguna acción / prospección (se asocia con el logro de una meta operando mediante simulación, predicción, intención o planificación), planificar qué hacer mediante el razonamiento sobre el estado actual del mundo o anticipando futuros estados (memoria episódica ⇒ explota recuerdos de experiencias pasadas, y memoria semántica ⇒ conocimiento del mundo), anticiparse al resultado a medida que ejecuta las acciones del propio robot o de otros agentes, personas u otros robots, aprender de la interacción resultante ya que las acciones futuras pueden ser más efectivas o eficientes (metacognición o meta-razonamiento) y adaptarse al cambio mediante el aprendizaje.
Razones de estudio: Construcción de robots inteligentes y entendimiento de la cognición.

[] TEMA 4: SISTEMAS COGNITIVOS (PARTE 2)

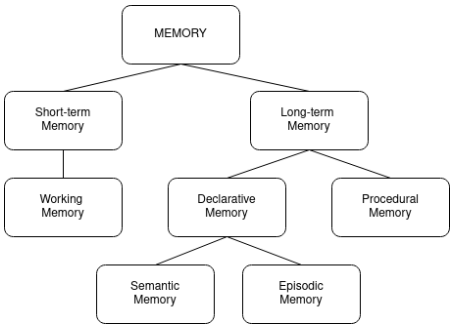
} 4.1 INTRODUCCIÓN

Memoria: Capacidad de almacenar, retener y recordar información (datos, experiencias, aprendizajes previos, entre otros). No sólo implica guardar datos, sino también organizarlos y hacerlos accesibles para su uso en procesos como el aprendizaje, la toma de decisiones o la imaginación. No es sólo un mecanismo pasivo encargado de almacenar conocimientos ocurridos en el pasado, también hace referencia al proceso cognitivo de guardar información que puede ser recuperada en el futuro, además de desempeñar un papel crucial y a veces inesperado en la cognición.

Conocimiento: Resultado de procesar y entender la información que se ha almacenado en la memoria, lo que implica una comprensión más profunda y la habilidad de aplicar la información a diferentes situaciones, por lo que no sólo se basa en recordar información, sino también en entenderla, relacionarla con otros datos y utilizarla de manera efectiva. Es fundamental en los sistemas cognitivos, ya que proporciona el contenido que complementa la arquitectura cognitiva.

Similitudes: Poseen un gran paralelismo y encapsulan la experiencia que surge de la interacción con el mundo.

Procesos: Recuperación de información, organización del conocimiento, reconocimiento / contextualización, aprendizaje y adaptación / toma de decisiones.



} 4.2 SHORT-TERM Y LONG-TERM

Short-term (a corto plazo / memoria de trabajo): Posee un breve almacenamiento y recuperación inmediata de detalles sustanciales.

Long-term (a largo plazo): Permite al ser humano sentirse contínuo y coherente en sus pensamientos, es decir, ser una persona continua con una vida continua, algo esencial de la interacción social entre las personas en la vida diaria, ya que permite recordar nombres, eventos, deberes, relaciones, etc.

} 4.3 DECLARATIVE Y PROCEDURAL

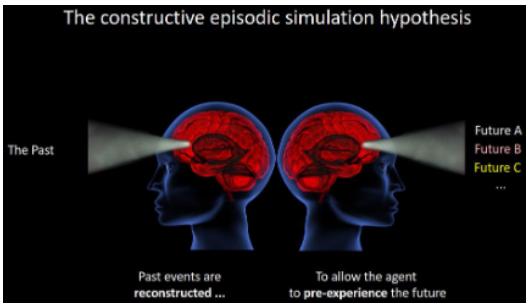
Declarative (“saber que”): Posee el conocimiento de las cosas / hechos, memoria proposicional (verdadero o falso) y explícita, puede comunicarse de un agente a otro a través del lenguaje y adquirirse en un único acto de percepción o cognición, pudiendo acceder al recuerdo consciente.

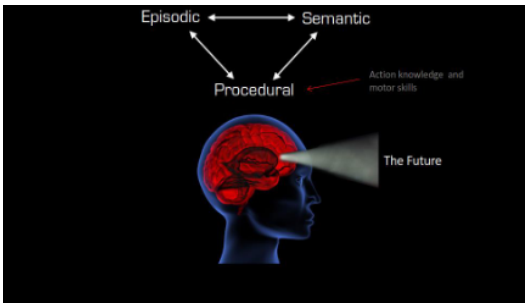
Procedural (“saber cómo”): Posee las memorias de las acciones orientadas a las habilidades, memoria implícita y no declarativa, ya que sólo puede demostrarse y adquirir progresivamente con cierta práctica, pero sin tener acceso a la memoria consciente.

} 4.4 EPISÓDICA Y SEMÁNTICA

Episódica: Consiste en instancias específicas de la experiencia del agente (autobiográficas) desde un contexto espacial y temporal explícito, siendo esta secuencia temporal el único elemento estructural de la memoria episódica. Esta clase de memoria es intrínsecamente constructiva, ya que los recuerdos episódicos antiguos se reconstruyen de forma ligeramente distinta cada vez que se asimila o recuerda un nuevo recuerdo episódico.

Hipótesis de la simulación episódica constructiva (Schacter y Addis, 2007): Permite la simulación de múltiples futuros posibles, lo que impone una necesidad aún mayor de capacidad constructiva debido a la necesidad de extrapolar más allá de las experiencias pasadas.





Semántica: Derivada de la memoria episódica a través de un proceso de generalización y consolidación, consiste en el conocimiento general sobre el mundo del agente (hechos, ideas y conceptos), siendo ésta independiente de la experiencias específicas del agente, ya que la memoria es necesaria para el uso del lenguaje.

} 4.5 MODAL Y AMODAL

Modal: Vinculada directamente a una modalidad sensorial concreta (visión, audio o tacto). Sin embargo, es más probable que la memoria episódica sea modal, ya que está estrechamente vinculada a las experiencias específicas de los agentes.

Amodal: No posee una asociación necesaria con las experiencias sensoriomotoras (hechos declarativos semánticos representados simbólicamente).

} 4.6 HETERO-ASSOCIATIVE Y AUTO-ASSOCIATIVE

Associative: Consiste en la vinculación de dos elementos de información o patrones, donde el primer elemento o patrón se utiliza para recordar al segundo.

Hetero-associative: Recupera un recuerdo de carácter diferente al de entrada (olor / sonido).

Auto-associative: Recupera un recuerdo de la misma modalidad que el que lo evocó (foto de un objeto).

} 4.7 CAPACIDAD COGNITIVA

Capacidad cognitiva: Simula internamente los resultados de posibles acciones seleccionando la más adecuada para la situación actual, ya que la memoria puede verse como un mecanismo que permite a un agente cognitivo prepararse para actuar, superando mediante la anticipación las limitaciones inherentes al “aquí y ahora” de sus capacidades perspectivas. Un sistema cognitivo no funciona sólo en función de los datos sensoriales que recibe, sino que se prepara para lo que espera, adaptándose a lo inesperado.

Recorrido hacia delante: Proporciona el elemento predicativo anticipatorio de la memoria, que sugiere una posible secuencia de acontecimientos que conducen a un objetivo deseado.

Recorrido hacia atrás: Permite explicar cómo pudo ocurrir un acontecimiento o imaginar cómo podría haber sido de otra manera.

} 4.8 FUNCIONES DE LA MEMORIA

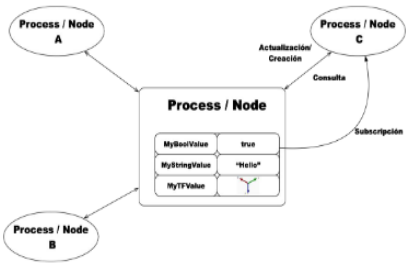
Funciones de cognición: Recordar acontecimientos pasados, anticipar acontecimientos futuros, imaginar el punto de vista de otras personas y navegar por el mundo.
Autoproyección: Capacidad de un agente para cambiar la perspectiva de sí mismo en el aquí y ahora mediante una simulación interna, que consiste en una construcción mental de una perspectiva alternativa imaginada.
Formas de simulación interna: Evocación de recuerdos episódicos (recuerdos del pasado), navegación (orientarse topográficamente relacionándose con el entorno actual), teoría de la mente (adoptar la perspectiva de otra persona sobre los asuntos) y prospección (anticipar posibles acontecimientos futuros).

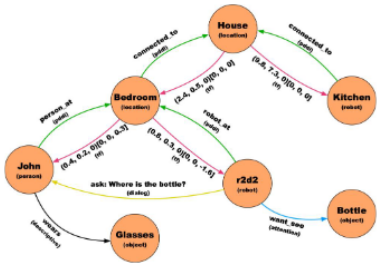
} 4.9 REPRESENTACIÓN DEL CONOCIMIENTO

Representación del conocimiento: Se ocupa de cómo el conocimiento puede definirse y manipularse de forma automatizada por programas de razonamiento.
Base de conocimiento: Colección de conocimientos representada utilizando el lenguaje de representación del conocimiento.
Sistema basado en conocimiento: Programa para ampliar y / o consultar una base de conocimiento.
Agente: Puede hacer proposiciones para determinar si estas expresiones son verdaderas o falsas a través del razonamiento y usar las expresiones como sus fuentes de conocimiento siempre que se pueda adquirir información y utilizar el razonamiento para determinar el resultado correcto.
Tipos de conocimiento según su representación: Conocimiento explícito (reglas y conceptos incorporados al agente) e implícito (infiere nuevos conceptos desconocidos para el agente).
Conocimiento de alto nivel: Consiste en una representación semántica y estructural de las relaciones entre distintos componentes.
Conocimiento de bajo nivel: Consiste en el programa de control y el sistema físico del robot, ya que no tienen estructura ni símbolos.

} 4.10 BLACKBOARD Y SISTEMAS BASADOS EN GRAFOS

Blackboard: Tabla clave/valor accesible desde los componentes de la arquitectura, donde el valor puede ser de cualquier tipo y puede implementarse como un objeto único en memoria o un nodo centralizado/distribuido.
Grafo: Establece que cada nodo del grafo es una instancia de un objeto, cuyas relaciones se codifican en los arcos. Éstos se diferencian en varios detalles como el tipo de nodos y arcos y su significado semántico, los cuales pueden ser distribuidos o centralizados.





} 4.11 SYMBOL GROUNDING PROBLEM

Puntos clave: Conexión entre símbolos y el mundo real, percepción / acción, interacción humana, aprendizaje / adaptación e integración multisensorial.
Representaciones icónicas: Permiten discriminar entre distintos objetos derivados directamente de datos sensoriales (imágenes visuales o patrones motrices).
Representaciones categóricas: Basadas en los resultados de procesos aprendidos e innatos que detectan características invariantes de categorías de objetos y acontecimientos a partir de los datos sensoriales (características de los objetos y comportamientos repetidos).

[] TEMA 5: ARQUITECTURAS COGNITIVAS (PARTE 1)

} 5.1 INTRODUCCIÓN

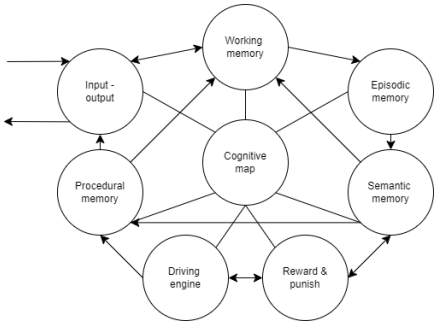
Hexágono cognitivo (Miller): Psicología, filosofía, lingüística, antropología, neurociencia e inteligencia artificial.
Arquitectura cognitiva: Framework software que integra todos los elementos requeridos por un sistema para que muestre las características de un agente cognitivo, cuyo diseño requiere la especificación de los formalismos para todos los procesos y representaciones del conocimiento utilizados por ese framework, y se centra en aquellos aspectos de la cognición que sean relativamente constantes en el tiempo e independientes de la tarea. Además, integra las capacidades cognitivas básicas (percepción, atención, memoria, aprendizaje, razonamiento, meta-razonamiento y prospección) para que puedan combinarse dinámicamente, lo que permite que el agente muestre un comportamiento flexible y sensible al contexto, seleccionando y controlando prospectivamente las acciones necesarias para alcanzar determinados objetivos. Y por último, debe ser capaz de desarrollarse de forma autónoma de modo que su rendimiento mejore con el tiempo y la experiencia.
Agente cognitivo: Posee la capacidad de actuar eficazmente en un mundo incierto, poco especificado, dinámico y que pueda operar con otros agentes cognitivos. Para ello, se requiere un sistema complejo que pueda construir modelos del funcionamiento del mundo, además de utilizarlos para guiar las acciones prospectivamente y actualizarlos dinámicamente a medida que el sistema aprende de sus interacciones.

} 5.2 TIPOS DE ARQUITECTURAS COGNITIVAS

Cognitivistas (Simbólicas): Se fundamentan en el uso de procesos computacionales definidos y estructurados, los cuales simulan comportamientos cognitivos mediante el uso de símbolos y reglas. Este enfoque intenta replicar la capacidad humana de procesar información de manera lógica y secuencial.
Emergentes: Se fundamentan en procesos de desarrollo y auto-organización donde las estructuras y habilidades cognitivas surgen gradualmente de la interacción contínua con el medio. Este grupo incluye a los sistemas conexionistas, que utilizan redes neuronales artificiales para simular cómo las interacciones entre neuronas pueden generar capacidades cognitivas complejas.
Híbridas: Integran elementos de ambos enfoques.

} 5.3 ARQUITECTURA COGNITIVA COGNITIVISTA

UTC (Teorías Unificadas de la Cognición): Abarcan una amplia gama de cuestiones cognitivas (atención, memoria, resolución de problemas, toma de decisiones y aprendizaje) desde varias áreas (psicología, neurociencia e informática). Desde el punto de vista cognitivista, una arquitectura cognitiva se centra en los aspectos de la cognición que son relativamente constantes en el tiempo e independientes de la tarea.
Modelo computacional genérico: No es específico del dominio ni de la tarea, cuyo conocimiento es habitualmente determinado por el diseñador, ya sea explícita o implícitamente, y es adaptado y aumentado por técnicas de machine learning (arquitectura cognitiva + conocimiento = modelo cognitivo).



Herencia: Solucionador general de problemas y la idea de regla de producción (condición y acción).
Objetivos: Conocer el funcionamiento de las reglas de producción, conocer la consecuencia de sus acciones, decidir qué regla utilizar en cada momento y, en ocasiones, resolver submetas hasta llegar a la meta principal.
Estado: Estructuras que describen la configuración particular dentro de la secuencia requerida para alcanzar un objetivo, compuesta por estados iniciales, intermedios y finales (estado inicial, intermedio y final).
Operador: Modifica un estado para convertirlo en otro mediante la aplicación de una regla específica, y es esencialmente la acción que lleva de un estado a otro dentro del proceso de resolución de problemas.
Espacio problema: Conjunto de estados y operadores disponibles para alcanzar el objetivo, donde las dimensiones y la complejidad del espacio dependen de la tarea.

} 5.4 ARQUITECTURA COGNITIVA EMERGENTE

Enfoques emergentes: Centrados en el desarrollo desde un estado primitivo a un estado plenamente cognitivo a lo largo de la vida del sistema. A diferencia del enfoque cognitivista, el término arquitectura cognitiva emergente, no es el framework que complementa el conocimiento sino que es el framework que facilita el desarrollo. Es todo lo que un sistema cognitivo necesita para ponerse en marcha.
Características: Tiene como base la ontogénesis (crecimiento, desarrollo, habilidades innatas y conocimientos básicos), una estructura en la que integrar diversos mecanismos (percepción, acción, adaptación, anticipación, motivación y el desarrollo de todas), se hace gran énfasis en la construcción de habilidades adaptativas, anticipatorias y que preserven la autonomía, teniendo en que cuenta la morfología del cuerpo físico en el que se inserta la arquitectura. Sin embargo, este enfoque rechaza el dualismo entre mente y cuerpo y el funcionalismo que trata los mecanismos cognitivos independientemente de la plataforma física (funcionalismo computacional o robótico).

} 5.5 CARACTERÍSTICAS DESEABLES

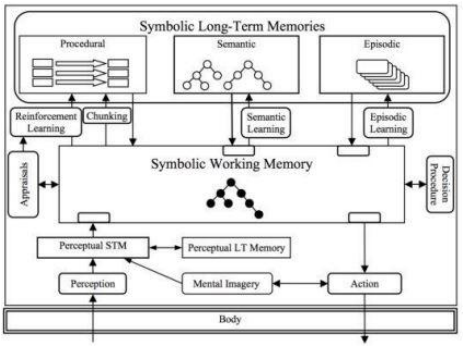
Realismo: Ecológico (capacidad de operar en entornos desconocidos, actividades cotidianas, incertidumbre y conflictos), bioevolutivo (la inteligencia humana se reduce a un modelo de inteligencia animal) y cognitivo (psicología humana, neurociencia y filosofía).
Características del comportamiento: Actuar / Reaccionar, esquema conceptual sencillo, ponderación simple de alternativas, secuencia temporal de acciones, comportamientos rutinarios aprendidos gradualmente y adaptación por ensayo y error.
Características cognitivas: Aprendizaje ascendente implícito (procesos inaccesibles e imprecisos), aprendizaje simbólico explícito (procesos accesibles y precisos), modularidad funcional física y enfoque híbrido de la cognición, donde los enfoques emergentes estrictos no cumplen el requisito de accesibilidad, en cambio, los enfoques cognitivistas sí que podrían cumplirlo.
Características deseables (Sun, 2007): Percepción, categorización, representaciones múltiples, múltiples tipos de memoria, toma de decisiones, razonamiento, planificación, resolución de problemas, metacognición, comunicación, control y ejecución de acciones y diversos tipos de aprendizaje.
Características deseables (Langley, 2009): Reconocimiento y categorización, toma de decisiones y elección, percepción y evaluación de situaciones, predicción y control, resolución de problemas y planificación, razonamiento y mantenimiento de creencias, ejecución y acción, interacción y comunicación, memoria, meditación y aprendizaje.

} 5.6 CAPACIDADES COGNITIVAS BÁSICAS

Percepción: Proceso encargado de transformar la información de entrada en una representación interna propia del sistema, haciendo uso de las distintas modalidades sensoriales (visión, audición y háptica).
Atención: Reduce la información que un sistema cognitivo tiene que procesar seleccionando la información relevante y filtrando la irrelevante utilizando mecanismos selectivos (elegir una entidad entre muchas), restrictivos (elegir de algunas entidades entre muchas) y supresores (suprimir algunas entidades de entre muchas es decir, características, objetos o lugares que no son relevantes).
Selección de acciones: Determina lo que el agente debe hacer a continuación, ya sea una planificación (determina una secuencia de pasos para evaluar un objetivo determinado antes de la ejecución del plan) o una selección dinámica de acciones (selección de una acción basada en el conocimiento del momento, normalmente utilizando mecanismos de selección de orden ganador, probabilístico o predefinido).
Memoria: Sensorial a corto plazo (percepciones recientes), de trabajo a corto plazo (relevante para la tarea actual), episódica a largo plazo (clave para la anticipación, autobiográfica), semántica a largo plazo (conocimiento general del mundo), procedimental a largo plazo (habilidades motrices) y global a largo plazo (para arquitecturas que no distinguen entre tipo y duración).
Aprendizaje: Capacidad de un sistema para mejorar su rendimiento a lo largo del tiempo mediante la adquisición de conocimientos o habilidades, ya sea de forma declarativa (adquisición de conocimientos explícitos), no declarativa (perceptivo, procedimental, asociativo y no asociativo), supervisada, no supervisada o por refuerzo.
Razonamiento: Capacidad de procesar lógica y sistemáticamente el conocimiento, normalmente para inferir conclusiones. Existen tres formas de inferencia lógica (deducción, inducción y abducción), donde el razonamiento se centra en el objetivo práctico de encontrar la siguiente (mejor) acción a realizar.
Metacognición: Capacidad que tiene un sistema cognitivo de controlar sus procesos cognitivos internos, razonar sobre ellos y adaptarlos. Es necesaria para la cognición social si el agente quiere formarse una teoría de la mente (toma de perspectiva), es decir, la capacidad de inferir los estados cognitivos de otros agentes con los que interactúa.
Prospección: Capacidad de anticipar el futuro, la cual se encuentra en el corazón de las otras características fundamentales de un agente cognitivo (autonomía, percepción, acción, aprendizaje y adaptación). Es fundamental para la acción, ya que las acciones están orientadas a objetivos y guiadas por información prospectiva, donde la simulación interna desempeña un papel clave.

} 5.7 ARQUITECTURA SOAR

Arquitectura SOAR: Arquitectura cognitiva híbrida con un sistema basado en reglas (operaciones if-then-else).
Características: Ciclo de producción (activa todas las reglas que coinciden con la información en la memoria de trabajo simbólica, actualiza la memoria y activa todas las reglas), ciclo de decisión (selecciona una acción), subobjetivo universal (crea un nuevo objeto y expone más conocimientos cuando se encuentra bloqueado) y aprende una nueva regla cuando se resuelve un bloqueo.
Memoria de trabajo: Almacena la situación actual, que incluye datos provenientes de sensores, inferencias intermedias, objetivos que están en curso y operadores activos, se estructura en objetos descritos a través de sus atributos, que pueden estar asociados a sub-objetos y exhiben una organización jerárquica.
Memoria de producción: Conserva el conocimiento procedimental de largo plazo, el cual define cómo actuar frente a distintas situaciones presentes en la memoria de trabajo.
Programa SOAR: Contiene el conocimiento que debe utilizarse para resolver una tarea específica (o un conjunto de tareas), información sobre cómo seleccionar y aplicar operadores para transformar los estados del problema y un medio para saber que se ha alcanzado el objetivo.



} 5.8 ARQUITECTURA ISAC

ISAC (Intelligent Soft Arm Control): Arquitectura cognitiva híbrida para un robot humanoide que consta de una colección integrada de agentes de software y memorias asociadas, donde los agentes funcionan de forma asíncrona y se comunican entre sí mediante el paso de mensajes. Además, comprende agentes activadores para el control del movimiento, perceptivos y de respuesta de primer orden (First-Order Respondent Agent FRA) para efectuar el control reactivo percepción-acción.

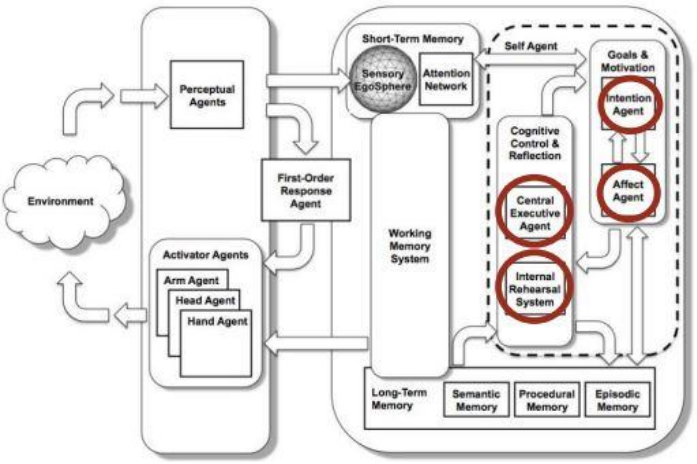
Short-Term Memory: Memoria espaciotemporal centrada en el robot de los acontecimientos perceptivos actuales denominada egoesfera sensorial (representación discreta de lo que ocurre alrededor del robot por una esfera geodésica indexada por dos ángulos). También dispone de una red atencional que determina los eventos perceptivos más relevantes.

Long-Term Memory: Almacena información sobre las habilidades aprendidas y las experiencias pasadas del robot (memoria semántica, episódica y procedimental).

Episodic Memory: Resume experiencias pasadas y crea vínculos o asociaciones entre ellas (situación exterior, objetivos, emociones, acciones, resultados que se derivan de las acciones y valoraciones de estos resultados), donde los episodios están conectados por enlaces que encapsulan comportamientos como transiciones de un episodio a otro en varios niveles.

Working Memory System: Almacena temporalmente información relacionada con la tarea que se está ejecutando en ese momento y encapsula expectativas de recompensa futura aprendidas mediante una red neuronal, donde se tiene un tipo de memoria caché para STM y la información que almacena (chunks).

Comportamiento cognitivo: Se consigue mediante la interacción de varios agentes con un subsistema de control cognitivo y reflexión compuesto por un agente ejecutivo central ACE (responsable del control cognitivo) y un sistema de ensayo interno (encargado de invocar habilidades para las tareas asignadas en el actual foco de atención y experiencias pasadas) y un subsistema de objetivos y motivación (un agente de intención encargado de proporcionar los objetivos y un agente de afecto modula la toma de decisiones).



Proceso: El First-Order Response Agent (FRA) produce respuestas reactivas a los estímulos sensoriales, además de ser el responsable de ejecutar las tareas. Cuando un ser humano le asigna una tarea, el FRA recupera la habilidad de la memoria procedimental en la LTM que corresponde a la habilidad descrita en la información de la tarea. A continuación, lo coloca en el WMS en forma de trozos junto con el concepto actual y el Agente Activador lo ejecuta, suspendiendo la ejecución siempre que se requiera una respuesta reactiva. Si el FRA no encuentra ninguna competencia adecuada para la tarea, el Agente Ejecutivo Central toma el relevo, para así recordar experiencias y comportamientos pasados de la memoria episódica, que contienen información similar a la de la tarea actual. Por último, se selecciona un par conducta-precepto, basándose en la percepción actual en el SES, su relevancia y la probabilidad de ejecución con éxito determinada por la simulación interna, y se coloca en la memoria de trabajo, donde el Agente Activador ejecuta la acción.

[] TEMA 6: ARQUITECTURAS COGNITIVAS (PARTE 2)

} 6.1 INTRODUCCIÓN

Aplicaciones prácticas (categorías): Modelización del rendimiento humano (HPM), juegos / rompecabezas, robótica, experimentos psicológicos, procesamiento del lenguaje natural (NLP), interacción persona-robot y persona-ordenador (HRI / HCI), visión por ordenador, categorización / agrupación (clustering), agentes virtuales y otros.

Robótica: Comprende casi una cuarta parte de las aplicaciones de las arquitecturas cognitivas, donde se encuentra la navegación / evitación de obstáculos (entornos no estructurados en un vehículo autónomo, robot móvil y vehículo marino no tripulado), la búsqueda / manipulación de objetos (agarre de objetos blandos, diferentes tipos de agarre, adaptación del agarre a latas / cajas de diferentes tamaños y una regla iCub), la implementación por pocas arquitecturas de múltiples habilidades para escenarios complejos (vendedor robótico, tutoría, evaluación médica y RoboCog) y las arquitecturas de motivación biológica centradas en el aspecto evolutivo de las habilidades físicas y la reconstrucción sensoriomotora (plataforma robótica infantil iCub que explora la adquisición de habilidades de locomoción / agarre / manipulación, Robots Dav y SAIL que aprenden navegación guiada por visión e ISAC que aprende affordances de agarre).

Interacción humano-robot / computador (HRI): Campo multidisciplinar que estudia diversos aspectos de la comunicación entre procesos y robots en el contexto de la robótica social, asistencial o del desarrollo. Dependiendo del nivel de autonomía que demuestre el robot, las interacciones van desde el control directo (teleoperación) hasta la plena autonomía del robot, lo que permite la colaboración entre iguales. Las arquitecturas cognitivas en el campo de la HRI son fundamentales para desarrollar robots que puedan interactuar de manera efectiva con los humanos en una variedad de entornos y situaciones (procesar y entender las señales humanas, tomar decisiones y aprender de las interacciones).

Objetivos: El robot debe reconocer, comprender y participar en situaciones de comunicación, ya sea de forma explícita (el humano se dirige verbalmente al robot) o implícita (el humano señala un objeto), además de ser capaz de participar en acciones conjuntas de forma proactiva (planificando y proponiendo los planes resultantes al humano) o reactiva (respondiendo adecuadamente a las acciones e indicaciones humanas), además de moverse y actuar de forma segura, eficiente y legible, teniendo en cuenta las reglas sociales como la proxémica.

Elementos: Comunicación, ejecución y acción conjunta. Ésta última parte de un objetivo conjunto previamente establecido y acordado a través del diálogo, un entorno físico estimado a través de las capacidades de detección exteroceptivas del robot y aumentado por inferencias extraídas de observaciones previas, y por último, un estado de creencias que incluye conocimientos de sentido común a priori y modelos mentales de cada uno de los agentes implicados (el robot y sus compañeros humanos).

NLP (Natural Language Processing): Campo multidisciplinar que estudia la comprensión del lenguaje escrito o hablado, desde la percepción auditiva de bajo nivel, el análisis sintáctico y la semántica, hasta la conversión en dominios limitados (primeros modelos de traducción automática en 1950).

Era de los sistemas basados en reglas (1960-1970): Grandes gramáticas escritas a mano por lingüistas y programadores, redes semánticas y marcos (frames) para representar el conocimiento del mundo.

Problemas: Altísimo coste de construcción y mantenimiento de reglas y dificultad para manejar ambigüedad, variabilidad y dominio abierto.

ELIZA (Weizenbaum, 1966): Imitaba el estilo de un terapeuta centrado en el paciente, donde ante cada entrada del usuario, ELIZA aplicaba transformaciones de texto y devolvía preguntas o reformulaciones que invitaban al interlocutor a profundizar en sus propios pensamientos.

Transición a métodos estadísticos y de aprendizaje (años 1980–2000): Modelos de n-gramas y corpus etiquetados, e introducción de técnicas de aprendizaje supervisado.

Revolución del Deep Learning y LLMs (desde 2018): Modelos preentrenados de gran escala, la arquitectura Transformer (Vaswani et al., 2017) como base de GPT de OpenAI (generación autoregresiva de texto) y BERT de Google (codificación bidireccional para comprensión), LLaMA (Meta), PaLM (Google), ChatGPT, etc, con miles de millones de parámetros.

Desafíos: Explicabilidad de las decisiones, alucinaciones y sesgos en los datos de entrenamiento.

} 6.2 INTERACCIÓN HUMANO-ROBOT

Tipo de modelo	Contenido principal	Ejemplo de uso
Modelo del usuario	Preferencias, habilidades o intenciones	Ajustar velocidad de trabajo según experiencia del operario
Modelo del entorno	Mapas 3D, objetos y zonas de seguridad	Planificar trayectorias que eviten colisiones
Modelo de sí mismo	Estado de baterías, fuerzas máximas y precisión de sensores	Decidir cuándo recargar o recalibrar

Interacción cognitiva humano-robot (HRI): Campo de investigación que pretende mejorar las interacciones entre los robots y sus usuarios mediante el desarrollo de modelos cognitivos para robots, que permitan comprender y procesar información de manera similar a como lo hacen los humanos.
Modelos mentales del robot: Seguimiento de las reglas sociales establecidas en las interacciones humano-robot, los cuales se interpretan de forma social por los humanos. Para ello, es necesario comprender la cognición social humana para desarrollar plataformas robóticas que se adapten a las expectativas y comportamientos de los usuarios y qué modelos mentales usan las personas para interpretar el comportamiento de los robots.
Modelos atribuidos al robot: Enfoque “científico-exploratorio” o mechanistic stance (la persona interpreta al robot como un objeto técnico sin emociones ni vida interior, se centra en sus partes, su programación y sus causas mecánicas, como por ejemplo, por qué gira la cabeza o qué sensor ha activado, manteniendo “distancia emocional”) y enfoque “relacional-animista” o relational-animistic approach (la persona atribuye al robot de cualidades de ser vivo como emociones, intenciones y necesidades, se antropomorfiza al robot de forma deliberada, se trata al robot como un bebé o una mascota, respondiendo de forma afectiva incluso cuando verbalmente afirma “sé que no es un ser vivo”.

} 6.3 ANTROPOMORFIZACIÓN

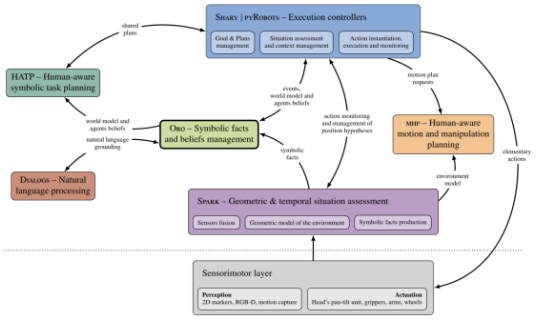
Antropomorfización: Proceso mediante el cual atribuimos características humanas a objetos no humanos (animales, objetos inanimados o robots). En el contexto de la interacción humano-robot, ésta puede ocurrir de varias formas y tener distintos efectos (diseño físico, comportamiento e interacción lingüística).
Mayor aceptación: Ocurre cuando los robots parecen y se comportan de manera más humana.
Empatía: Ocurre cuando los robots están diseñados para parecer vulnerables o expresar emociones.
Expectativas exageradas: Ocurre cuando las personas atribuyen cualidades humanas a los robots que no poseen.
Mejora de la comunicación: Ocurre cuando se limita el comportamiento humano.
Embodiment físico: Un robot físico despierta más proyección de vida e intenciones que un agente puramente virtual.
Co-presencia y entorno compartido: Cuando el robot está “a mi lado” o comparte mi mismo espacio (oficina, salón, taller), siento que sus acciones tienen relevancia social directa.
Normas de cortesía y etiqueta: Robots que saludan, esperan turno para hablar o respetan el espacio personal refuerzan la idea de “compañero social” y no solo “herramienta”.

} 6.4 TEORÍA DEL VALLE INQUIETANTE

Teoría del valle inquietante: Sugiere una conexión entre la apariencia de un objeto y la reacción emocional hacia éste, basándose en su similitud con un ser humano. Sin embargo, cuando los robots humanoides se parecen demasiado en aspecto y comportamiento a un ser humano real, se genera una sensación de repulsión en las personas que los observan.

} 6.5 INTERACCIÓN HUMANO-MÁQUINA

Modelo basado en diálogos: Establece un modelo común en actividades conjuntas entre humanos y robots, los cuales buscan facilitar la comprensión mutua y mejorar la colaboración entre tareas específicas (navegación y exploración colaborativa) para compartir información y tomar decisiones conjuntas sobre el control en puntos clave de la tarea. También se pueden abordar las tareas del dominio y del diálogo como acción conjunta, donde se integran módulos de interpretación semántica y toma de decisiones centralizada, utilizando recursos como la historia del discurso en curso, un modelo del mundo y un plan de dominio para generar comportamientos de acción y comunicación.



ORO (Base de Conocimiento): Almacena el conocimiento simbólico, y maneja y procesa el conocimiento de forma simbólica y lógica.

SPARK (Razonamiento espacial y la evaluación del entorno): Realiza evaluación de la situación física en el entorno del robot, lo que le permite generar conocimiento simbólico sobre la relación espacial entre los objetos y los agentes (humanos y robots).

HATP (Planificador Simbólico de Tareas): Decide qué acciones debe realizar el robot para alcanzar un objetivo dado basándose en un conjunto de reglas y hechos, produce planes colaborativos (sincronización) y se asegura que las tareas cumplan las condiciones de éxito de las acciones planeadas.

DIALOGS (Procesamiento de Diálogo y Comunicación Multimodal): Procesamiento de lenguaje natural y la interacción verbal entre el robot y los humanos, resuelve ambigüedades, entiende las intenciones del humano, y usa el conocimiento en la base de datos para interpretar y generar respuestas apropiadas.

SHARY (Controlador de ejecución): Gestiona la ejecución de planes y coordinación de acciones, ejecuta las acciones físicas del robot y supervisa su progreso, y controla la ejecución de planes colaborativos elaborados por HATP, asegurando que se coordinen las acciones entre el robot y los humanos.

MHP (Human-Aware Motion Planning): Complementa el control de ejecución en la arquitectura del robot, planifica los movimientos geométricos del robot de manera que no solo sean eficientes y seguros, sino también socialmente aceptables para los humanos con los que interactúa.

Acción conjunta en HRI: Meta común (establecida y acordada previamente normalmente vía diálogo), entorno físico (percibido por los sensores exteroceptivos del robot y enriquecido con inferencias de observaciones previas) y estado de creencias (conocimiento a priori / sentido común y modelos mentales compartidos de todos los agentes tanto en robots como humanos).

Ejemplos: Desambiguación implícita a través de la perspectiva visual y desambiguación explícita a través de la interacción verbal y gestos.

Planificación de acciones: Ejecuta un plan compartido a partir de un planificador simbólico de tareas y un controlador de ejecución.

[] TEMA 7: ARQUITECTURAS COGNITIVAS (PARTE 3)

} 7.1 INTRODUCCIÓN

Lenguaje (herramienta cognitiva): Permite organizar experiencias y representar información abstracta mediante la formación de conceptos, facilita el pensamiento más allá de lo sensorial inmediato apoyando el razonamiento abstracto y la resolución de problemas, modula procesos de codificación, almacenamiento y recuperación de información, y desarrolla habilidades sociales.
Tipos: Representaciones simbólicas clásicas (lógica, reglas de producción, redes semánticas y frames), transición conexionista (redes neuronales simples y modelos recurrentes como RNN o LSTM), incorporación de la probabilidad (modelos estadísticos de lenguaje como N-gramas, Modelos Ocultos de Markov HMM, Redes Bayesianas y Modelos Gráficos Probabilísticos CRF) y las redes de atención y Transformers (atención y auto-atención).

} 7.2 LARGE LANGUAGE MODELS

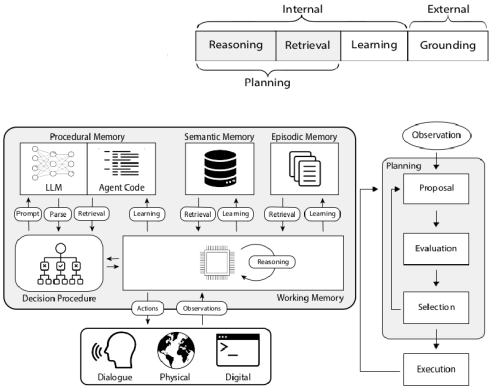
LLM: Transformer models entrenados para predecir la siguiente palabra (token) entrenados con millones de sentencias (textos de internet). Si se muestrean múltiples tokens consecutivos, sería posible imitar conversaciones y usar el LLM para generar respuestas más detalladas. Sin embargo, si se continúa la conversación, cualquier LLM mostrará su principal limitación sin recordar conversaciones previas. Y por último, para poder resolver cualquier tipo de limitación, las herramientas basadas en LLM han incorporado mecanismos para realizar un proceso de razonamiento (división del problema en pasos más pequeños) antes de dar la respuesta.
Enfoques híbridos neuro-simbólicos: Razonamiento basado en grafos (gráficos de conocimiento y LLMs para la toma de decisiones estructurada), razonamiento aumentado con herramientas (LLMs que invocan APIs, ejecutan código o consultan bases de datos) y aprendizaje por refuerzo a partir de retroalimentación humana RLHF (mejora la coherencia humana).
Chain-of-Thought (CoT) Prompting: Fomenta que los modelos generen pasos intermedios antes de llegar a una conclusión. Este método mejora la capacidad de los modelos para resolver tareas complejas estructurando el proceso de resolución de problemas.
Tree-of-Thought (ToT) Prompting: Permite a los modelos explorar múltiples rutas de razonamiento en una estructura de árbol, lo que facilita la revisión de estrategias y mejora la flexibilidad para resolver problemas.
Self-Consistency Decoding: Implica generar varias rutas de razonamiento de manera independiente y seleccionar la respuesta más consistente, ayudando a reducir sesgos y aumentando la fiabilidad de las respuestas del modelo.
Toolformer (Language Models Can Teach Themselves to Use Tools): Muestra cómo los LLMs pueden aprender a usar herramientas externas para mejorar su capacidad de razonamiento.
Reflexión (Language Agents with Verbal Reinforcement Learning): Introduce el marco de reflexión que permite a los modelos mejorar sus respuestas mediante el aprendizaje por refuerzo verbal.

} 7.3 LANGUAGE AGENTS

LLMs (Large Language Models): No pueden realizar por sí mismos tareas que requieran una toma de decisiones compleja ni ejecutar acciones del mundo real, donde el mundo puede ser complejo (muchas reglas que manejar) o estocástico (las reglas disponibles pueden fallar).
Agente inteligente: Entidad capaz de percibir su entorno, procesar percepciones y responder o actuar en dicho entorno de manera racional, logrando objetivos que tienden a maximizar un resultado esperado y adquiriendo conocimiento a su desempeño. Además, es capaz de percibir su medio ambiente con la ayuda de sensores y actuadores (elementos que reaccionan a un estímulo realizando una acción).
Niveles: Nivel 1 Agente de Texto (utiliza la acción de texto y la observación como ELIZA y LSTM-DQN), Nivel 2 Agente LLM (utiliza LLM para actuar como SayCan y Language Planner) y Nivel 3 Agente de Razonamiento (utiliza LLM para actuar como ReAct y AutoGPT).
RL tradicional: Aprendizaje con recompensa escalar y mediante actualización de pesos.
RL verbal: Aprendizaje con retroalimentación de texto y mediante actualización del lenguaje (memoria a largo plazo).
Language Agents: Colocan al LLM en un bucle de retroalimentación directa con el entorno externo transformando las observaciones en texto y utilizando el LLM para elegir acciones.

} 7.4 ARQUITECTURAS COGNITIVAS PARA LANGUAGE AGENTS

Cognitive Language Agents: Simulan una arquitectura cognitiva utilizando LLMs para gestionar el estado interno del agente a través de procesos como el aprendizaje y el razonamiento.
COgnitive Architectures for Language Antes (CoALA): Framework conceptual para caracterizar y diseñar agentes lingüísticos de propósito general (almacenamiento de información en memoria a corto / largo plazo, un espacio de acciones internas y externas, y un procedimiento de toma de decisiones basado en planificación y ejecución).
Memoria de trabajo a corto plazo: Posee información relevante (variables simbólicas) para el ciclo de decisión actual, además de un conjunto de datos que persiste a través de las llamadas LLM que pueden ser usadas para razonar, y por último, es capaz de conectar diferentes componentes, como la memoria a largo plazo y los entornos externos.
Memoria semántica a largo plazo: Basada en el conocimiento de hechos sobre él mismo y el mundo mediante el uso de fuentes de datos externas para aumentar el conocimiento del LLM, que pueden actualizarse para mejorar esta memoria.
Memoria episódica a largo plazo: Compuesta por el historial de experiencias del agente en ciclos de decisión anteriores, cuyo contenido se recupera en la memoria de trabajo y se utiliza para razonar y tomar decisiones.
Memoria procedimental a largo plazo: Posee un conocimiento implícito (integrado en los pesos del LLM) y explícito (escrito por el desarrollador en el código del agente, capaz de implementar acciones y procedimientos que facilitan la toma de decisiones), la cual se inicializa cuando el agente se pone en marcha.



} 7.5 ACCIONES

Acciones: Externas (interacción con entornos externos y retroalimentación en la memoria de trabajo) e internas (interacción con la memoria interna y actualización de la memoria semántica, episódica y procedimental).
Grounding: Ejecutan acciones externas y procesan la información del entorno en la memoria de trabajo en forma de texto, lo que simplifica de forma efectiva la interacción del agente con el mundo exterior como un juego de texto con observaciones y acciones textuales (entornos físicos, diálogo con humanos u otros agentes y entornos digitales).
Recuperación: Lee información de las memorias a largo plazo a la memoria de trabajo, la cual podría implementarse de varias maneras dependiendo de la información y del tipo de memoria.
Razonamiento: Procesa, lee y escribe el contenido de la memoria de trabajo para generar nueva información, lo que permite al agente resumir y extraer información sobre la observación/trayectoria más reciente o la información recuperada de la memoria a largo plazo. Además, puede utilizarse para apoyar el aprendizaje o la toma de decisiones.
Aprendizaje: Ocurre escribiendo información en la memoria a largo plazo mediante la actualización de la memoria episódica con la experiencia, la memoria semántica con el conocimiento, los parámetros LLM y el código del agente.
Toma de decisiones: Ciclo formado por una planificación (razonamiento/recuperación para proponer y evaluar acciones) y una ejecución (selecciona una acción de aprendizaje/ground para afectar a la memoria interna o al mundo externo).