

Ingeniería y Gestión del Conocimiento: práctica 1

Miguel Expósito Martín - 72056097H

Introducción

El presente documento sintetiza las decisiones tomadas a la hora de llevar a cabo la presente práctica. Cabe destacar que, como autor de este material, es la primera vez que mantengo un contacto con un sistema basado en reglas en general y con el entorno CLIPS en particular.

Entregables

- Directorio /md : código fuente de la memoria en [Markdown](#).
- Directorio /png : imágenes asociadas a los diagramas.
- Directorio /uml: código fuente de los diagramas en [PlantUML](#).
- Directorio /ods: hoja de cálculo con alimentos en [Libreoffice](#).
- Archivo hechos_iniciales.clp: plantillas y hechos iniciales necesarios para la ejecución de las reglas.
- Archivo dieta.clp: reglas que resuelven el problema en base a los hechos iniciales.

Entorno y ejecución

Dada la similitud de la sintaxis de CLIPS con LISP, se optó por utilizar un Entorno Integrado de Desarrollo (IDE) para aprovechar funcionalidades como el resaltado de sintaxis. En concreto, se ha utilizado [IntelliJ IDEA](#) con un plugin de Clojure, cuya sintaxis es similar.

El código se ha ejecutado con éxito tanto en entornos Windows como en entornos Linux. Para replicar su ejecución, y suponiendo que el ejecutable de clips esté en la variable de entorno \$PATH del sistema correspondiente, desde el directorio raíz de la práctica en el que se encuentran todos los fuentes, hay que ejecutar las siguientes acciones **en orden**:

```
clips
(load ./hechos_iniciales.clp)
(load ./dieta.clp)
(reset)
(run)
```

Es decir, una vez en el entorno de CLIPS, primero se cargan las plantillas y hechos iniciales y posteriormente las reglas. Con el comando `reset()` se inicializa la base de datos de hechos y las reglas quedan listas para ejecutarse.

Enfoque

La planificación del desarrollo de la práctica se ha dividido en cuatro grandes bloques:

- Cálculo de los gastos energéticos y requerimiento calórico diario.
- Generación de combinaciones de alimentos (dietas) diarias.
- Reparto de alimentos de cada dieta en 5 comidas.
- Salida y entrada de datos.

La primera parte no revistió prácticamente dificultad, pudiéndose programar los cálculos en un tiempo razonable. En cambio, la generación de combinaciones de alimentos ha requerido de mucho tiempo y esfuerzo. La tercera parte presentó una dificultad media y la última fue muy sencilla.

Modelo de conocimiento

El modelo utilizado para la práctica se ha simplificado al máximo de cara a minimizar cualquier complejidad derivada de su estructura a lo largo del desarrollo de la misma. Básicamente consta de las siguientes entidades principales:

- Persona: contiene los hechos asociados a una persona, introducidos por teclado.
- Alimento: contiene los hechos asociados a los distintos alimentos (calorías, proteínas, etc.); los datos se han obtenido de diversas fuentes en Internet Ver referencia. Todos los datos de macronutrientes y calorías están referidos a 100g de alimentos. Como parte de la información relativa a alimentos se encuentra el campo **ración**. Este **se ha completado con las cantidades normales para una dieta para adultos de en torno a 2000 kcal diarias**, con datos extraídos de diversos artículos sobre nutrición. A partir de ahí, se han añadido raciones mayores y menores con objeto de ofrecer más flexibilidad a la hora de seleccionar alimentos/raciones necesarios para la dieta.
- Dieta: contiene los hechos asociados a una dieta: sus calorías totales, proteínas totales, etc. así como los alimentos asociados. Las entidades de tipo dieta son generadas por el sistema, con lo que se trata de conocimiento adquirido a lo largo de la ejecución del programa.

Al margen de estas entidades, se tienen las derivadas de la dieta (desayuno, almuerzo, etc.) y otras auxiliares utilizadas fundamentalmente para llevar un control del día cuya dieta se está generando, alimentos repetidos entre días, etc.

En el siguiente diagrama se clarifica la explicación anterior:

Plantillas

Se han definido las siguientes plantillas:

- Alimento
- Persona
- Dieta
- Desayuno-dieta
- Almuerzo-dieta
- Comida-dieta
- Merienda-dieta
- Cena-dieta
- Usados
- Usados-reparto
- Dia

Hechos iniciales

Junto con las plantillas están definidas las cargas de los siguientes hechos iniciales:

- Introducir-auxiliares
- Introducir-comidas
- Introducir-alimentos

Para generar los hechos asociados a los alimentos se ha utilizado una hoja de cálculo Libreoffice (incluida entre los entregables). Utilizando funciones de concatenación y sustitución, a partir de los datos sobre los alimentos distribuidos en columnas, se generan los hechos con la sintaxis correcta.

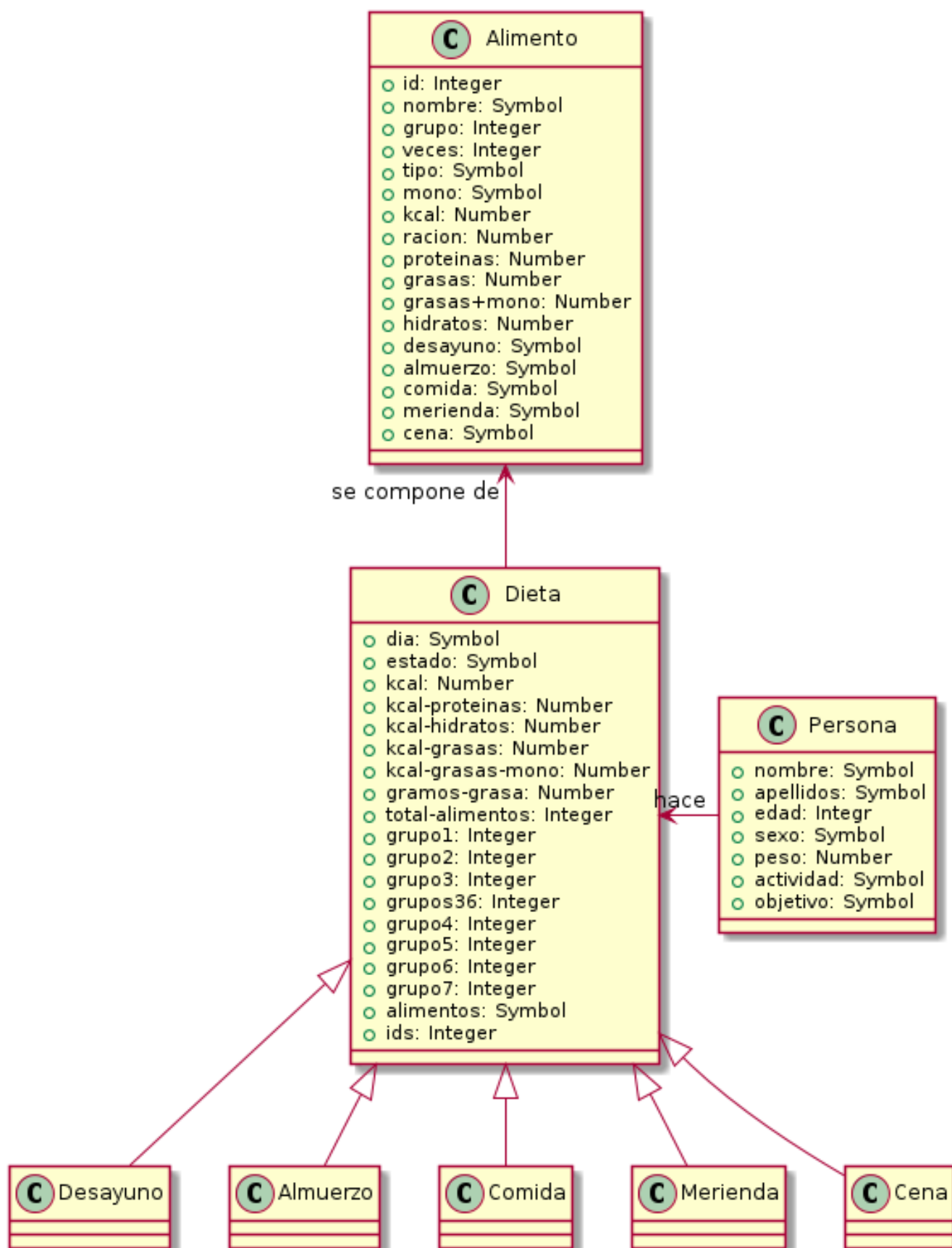


Figure 1: Modelo de conocimiento

Restricciones impuestas

- La suma total de calorías de la dieta debe ser inferior al requerimiento calórico diario.
- El requerimiento calórico diario se calcula como la suma del gasto energético basal y el gasto energético por actividad diaria.
- Los alimentos deberán repartirse en las siguientes raciones por grupos de alimentos:
 - 4-6 raciones/día de alimentos de los grupos 3 y 6
 - 2-4 raciones/día del grupo 4
 - 2-3 raciones del grupo 5
 - 2-3 raciones del grupo 1
 - 2-3 raciones del grupo 2
 - 40-60 gramos de grasa
- Las dietas que se generen tendrán que ser variadas.
- La contribución porcentual de macro-nutrientes a las calorías totales deberá ser aproximadamente de:
 - 50 - 55% carbohidratos
 - 30 - 35% grasas. (de las cuales, 15 - 20% monoinsaturados)
 - 10 - 15% proteínas.

Sistema de reglas

Cálculo de gastos energéticos y requerimiento calórico diario

Esta fase no presentó mayores dificultades. Básicamente se han traducido los datos proporcionados en la tabla del enunciado por reglas que calculan los distintos valores en función de las edades y el sexo de los individuos. El enfoque ha consistido en crear una regla por cada combinación sexo-edad para calcular el gasto energético basal. Posteriormente, se calcula el gasto debido a la actividad en función de los datos introducidos y finalmente el requerimiento calórico diario con simples reglas.

Primeros pasos en la generación de dietas

En primer lugar, se trató de plantear el problema siguiendo como modelo el número 11 de la colección de ejercicios resueltos (composición de programas musicales en emisoras de radio). Debido a mi conocimiento, formación y experiencia en lenguajes procedimentales y orientados a objetos, este enfoque resultó en un sesgo hacia que las reglas se ejecutaran como si de uno de estos lenguajes se trataran, lo cual resultó contraproducente. Se generaron muchas reglas, muchas variables de control por fases, muchas excepciones a los casos generales... pero no se consiguió llegar a una sólo solución del problema (dieta que cumpliera los requisitos diarios).

A continuación, pensando más detenidamente en el funcionamiento de un sistema basado en reglas, se optó por plantear el problema como si de una sentencia declarativa (similar a SQL, por ejemplo) con condiciones se tratara. Es decir, se intentó que el sistema proporcionara combinaciones de todos los alimentos posibles que satisficieran las condiciones impuestas en cuanto a número de calorías y proporciones de proteínas, hidratos, etc. El problema contaba con un mínimo de doce variables o incógnitas, y al introduciras en una sola regla lo único que se conseguía era bloquear el sistema totalmente (e incluso la máquina en la que se ejecutaba).

Algoritmo de generación de dietas

Finalmente, se trató de darle al problema el mismo enfoque que al “problema de la mochila”. En este tercer planteamiento se tomó como base el número 13 de la colección de ejercicios resueltos (extracción de billetes de un cajero automático). El símil sería sencillo: los billetes equivaldrían a alimentos y los reintegros a las dietas. En resumidas cuentas el problema se planteaba también como un espacio de estados, donde los nodos serían las dietas. Existiría un operador por cada grupo de alimentos y su selección incrementaría las calorías totales en función

de las raciones y los alimentos seleccionados. En el nodo final (o solución), la suma total de calorías equivaldría aproximadamente al requerimiento calórico diario con las restricciones impuestas.

Tan sólo restaba establecer el criterio de decisión para “cerrar” un nodo. En principio se trató de maximizar el número de calorías, pero las combinaciones de alimentos resultantes estaban muy alejadas de las requeridas. Posteriormente se estableció como criterio maximizar el número de alimentos elegidos hasta la fecha en la dieta, lo que proporcionaba soluciones esporádicamente y con un rendimiento no aceptable (decenas de minutos).

A la vista de los resultados, se trató de “guiar” al algoritmo obligando a seleccionar determinados alimentos por fases. En este punto se probaron diversos criterios: elegir primero los alimentos proteicos, las grasas... ninguno conseguía que el algoritmo encontrara soluciones en tiempos razonables. Por ello, se optó por consultar un manual de alimentación y nutrición y documentación diversa sobre la confección de dietas (referencias). En todos ellos se indicaba que **el primer grupo de alimentos que debe ser seleccionado es el de carbohidratos** y que, una vez alcanzado el requisito calórico de este macronutriente, se proporcionaría a su vez el 50% de las proteínas necesarias. Asimismo, en dicho manual se indica que se ha de seguir con verduras y frutas para terminar con lácteos y proteínas, quedando en último lugar las grasas. Este ha sido el criterio usado para guiar al algoritmo.

Generación de dietas

Las dietas se generan a través de los distintos operadores, uno para cada grupo de alimento. Estos operadores seleccionan todas las posibles dietas en estado **cerrado**, es decir, que ya han sido seleccionadas previamente como candidatas prometedoras, y añaden un alimento de su grupo a la dieta computando los totales y subtotales de calorías, proteínas, hidratos, etc.

La generación de las nuevas dietas se realiza de forma que el alimento no se haya seleccionado ya en el mismo día. También se tienen en cuenta las periodicidades semanales introducidas en la base de hechos (una, dos, tres y cinco veces por semana). Asimismo, se incorporan a la selección las restricciones impuestas en el enunciado con objeto de que el espacio de posibles candidatos sea menor y por tanto mejore la eficiencia del algoritmo.

Las reglas de los distintos operadores se lanzan cuando se ha calculado el requerimiento calórico diario y existe una entidad de tipo dieta.

Hay que tener en cuenta que, si bien es posible seleccionar entre distintas raciones para un mismo alimento, esta selección solo puede llevarse a cabo una vez en el mismo día. Es decir, la ración escogida será la definitiva; a partir de ahí, el alimento pasará a estar en la lista de utilizados y no será reutilizable en dicha dieta.

Selección de dietas

Una vez generados los posibles nodos candidatos con dietas, se utilizan los siguientes criterios para seleccionar uno prometedor:

- Primero, 2 alimentos del grupo 3. Con esto se pretende garantizar un mínimo de 2 raciones de legumbres semanales.
- Segundo, 4 alimentos de los grupos 3 o 6. La selección de todos los carbohidratos está restringida a determinadas proporciones de proteínas, hidratos y grasas, encuadrándolos dentro de unos límites. Dichos límites han sido seleccionados de forma heurística, si bien se ha utilizado el sentido común (por ejemplo, en la selección del tercer alimento se restringe el número de calorías de carbohidratos a la mitad del máximo, dado que se está seleccionando la mitad de los alimentos de dicho grupo).
- Tercero, 3 alimentos del grupo 4 (verduras). En este caso sólo era necesario prestar atención a los límites máximos de hidratos y proteínas.
- Cuarto, 2 alimentos del grupo 5 (frutas). Ídem que anterior.
- Quinto, 2 alimentos del grupo 1 (lácteos). En este caso se vigilan las proteínas.
- Sexto, 2 alimentos del grupo 2 (carne, huevos, pescado). Ídem que anterior.
- Séptimo, 1 ó 2 alimentos del grupo 7 (grasas), siempre comprobando que la cantidad de monoinsaturadas se encuentre dentro de los límites indicados en el enunciado.

Las dietas se seleccionan cambiando su estado a **cerrado**, para que sean explotadas en la siguiente iteración.

Diagrama de estados

A continuación se puede visualizar más cómodamente la descripción anterior en un diagrama de estados.

Identificación de la solución

Se trata una regla de alta prioridad (99) que comprueba si existen dietas que cumplan con los requisitos calóricos y de grupos de alimentos. No es necesario comprobar restricciones sobre los macronutrientes porque la forma de guiar al algoritmo garantiza que, una solución cercana al requisito calórico diario, los cumplirá.

En caso de encontrarla, cambia su estado a **solucion** y añade sus alimentos al listado de alimentos usados, para poder ser comprobados en dietas posteriores.

Otras consideraciones algorítmicas

Control de flujo

EL control de flujo se realiza en base a dos técnicas:

- Variables de control
- Prioridades de reglas

En el primer caso se utilizan variables como `?rcd` o `?dia` como disparadores para las reglas de selección y reparto de alimentos, respectivamente.

El segundo caso se utiliza, por ejemplo, en la distribución de alimentos; estos se van repartiendo en un orden modelado por las distintas reglas a través de sus prioridades.

- Cálculo de gastos energéticos y requerimiento calórico diario: se disparan cuando existe una entidad de tipo persona.
- Generación de alimentos para la dieta: se disparan cuando se ha calculado el requerimiento calórico diario y existe una dieta en estado cerrado.
- Selección de alimentos para la dieta: tienen prioridad -10 y se disparan cuando se han generado todas las dietas y estas están en estado abierto.
- Distribución de alimentos en cinco comidas: se disparan cuando hay dietas en estado **solucion**. Se van ejecutando secuencialmente a través de sus prioridades (de 102 a 88).
- Salida por pantalla: con prioridad -100, se disparan después de que se han ejecutado todas las reglas de distribución de alimentos.

Realización de dietas variadas

Para la realización de dietas variadas se ha incluido una columna en el modelo de alimentos llamada **veces**, que incluye el número de repeticiones semanales que puede darse para un alimento dado (1, 2, 3 y 5).

El criterio para asignar a los alimentos un número determinado de repeticiones se ha extraído de las distintas referencias sobre nutrición y dietas consultadas. Hay datos que son de sabiduría popular, como por ejemplo:

- El aceite de oliva se puede consumir a diario, y de hecho es recomendable
- Es recomendable comer legumbres al menos dos veces por semana
- Es recomendable pescado entre dos o tres veces por semana
- Las verduras pueden tomarse todos los días de la semana
- Etc.

Por otra parte, para que reducir repeticiones de alimentos entre un día y el siguiente, se han asociado los que se toman dos veces a la semana a los martes y jueves y los que se toman tres veces a la semana a los lunes, miércoles y viernes. Los alimentos que se van utilizando en todas las dietas se almacenan en un hecho específico.

Finalmente, en el mismo día no se pueden repetir alimentos (se realiza una comprobación en la selección de los mismos contra un hecho de tipo multislot que contiene los alimentos asociados a la dieta en cuestión).

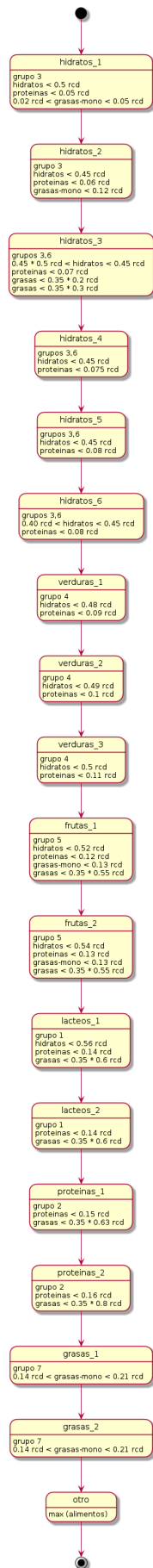


Figure 2: Diagrama de estados

Distribución de calorías en cinco comidas diarias

Para distribuir los alimentos de la dieta diaria en cinco comidas se han seguido los siguientes criterios extraídos de las referencias documentales:

Reparto de calorías:

- Desayuno: 30%
- Almuerzo y merienda: 10% cada uno
- Comida y cena: complementarios, tratando de ser la cena más ligera

Inclusión de alimentos según grupo con carácter de mínimo

- Desayuno: lácteos y carbohidratos
- Almuerzo y merienda: fruta
- Comida: proteínas, legumbres, carbohidratos, verduras y grasa
- Cena: resto

Márgenes de error

Las distribuciones de calorías y alimentos no son exactas. Esto se debe al carácter aproximado de la disciplina de la nutrición en general, tal y como se expone en el enunciado de la práctica, y de cara a la realización de las dietas se ha impulsado conscientemente. Por ejemplo:

- Los límites calóricos superiores están calculados frente al requisito calórico diario y se ha dejado cierta holgura (56% en vez de 55%, por ejemplo). Esto hace que, a la hora de calcular las proporciones en la dieta real, los porcentajes puedan ser ligeramente superiores sin suponer error o problema alguno.
- Algunos valores en la selección de alimentos no se han limitado conscientemente, para dejar cierta libertad al algoritmo.
- La distribución de alimentos en comidas no se ha llevado a cabo siguiendo unos porcentajes estrictos para poder encajar con mayor facilidad las combinaciones.

Otras consideraciones

En el foro de la asignatura se han comentado situaciones anómalas, como la incapacidad para diseñar dietas con 40-60 g de grasa en menores de tres años. Es evidente que, en lactantes, tanto la grasa como el resto de macronutrientes se aportan a través de la leche materna. En cualquier caso, los valores manejados en el enunciado de la práctica se refieren a sujetos adultos según la literatura Referencia. Estos porcentajes cambian para menores de tres años (ver Referencia). Es por ello que no se plantea acción alguna en el programa, que debería usarse con sentido común y sólo para dietas a partir de los 4 años.

Evaluación del funcionamiento del programa

Perfil 1: varón adulto de 83 kg y actividad moderada cuyo objetivo es mantener

Perfil 2: mujer adulta de 64 kg y actividad ligera cuyo objetivo es reducir

Perfil 3: varón adulto de 95 kg y actividad reposo cuyo objetivo es reducir

Perfil 4: niña de 18 kg y actividad moderada cuyo objetivo es mantener

Perfil 5: mujer de 60 kg y actividad intensa cuyo objetivo es mantener

Conclusiones y líneas de trabajo futuras

Referencias